METHODS FOR RAPID IDENTIFICATION OF PATHOGENS IN HUMANS AND ANIMALS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of U.S. application Serial No. 10/323,233 filed December 18, 2002, which is incorporated herein by reference in its entirety. This application is also a continuation-in-part of U.S application Serial No. 09/798,007 filed March 2, 2001, which is incorporated herein by reference in its entirety. The present application also claims priority to U.S. provisional application Serial No. 60/431,319 filed December 6, 2002, U.S. provisional application Serial No. 60/443,443 filed January 29, 2003, U.S. provisional application Serial No. 60/443,788 filed January 30, 2003, and U.S. provisional application Serial No. 60/447,529 filed February 14, 2003, each of which is incorporated herein by reference in its entirety.

STATEMENT OF GOVERNMENT SUPPORT

This invention was made with United States Government support under DARPA/SPO contract BAA00-09. The United States Government may have certain rights in the invention.

FIELD OF THE INVENTION

The present invention relates generally to clinical applications of directed to the identification of pathogens in biological samples from humans and animals. The present invention is also directed to the resolution of a plurality of etiologic agents present in samples obtained from humans and animals. The invention is further directed to the determination of detailed genetic information about such pathogens or etiologic agents.

The identification of the bioagent is important for determining a proper course of treatment and/or eradication of the bioagent in such cases as biological warfare and natural infections. Furthermore, the determination of the geographic origin of a selected bioagent

will facilitate the identification of potential criminal identity. The present invention also relates to methods for rapid detection and identification of bioagents from environmental, clinical or other samples. The methods provide for detection and characterization of a unique base composition signature (BCS) from any bioagent, including bacteria and viruses. The unique BCS is used to rapidly identify the bioagent.

BACKGROUND OF THE INVENTION

In the United States, hospitals report well over 5 million cases of recognized infectious disease-related illnesses annually. Significantly greater numbers remain undetected, both in the inpatient and community setting, resulting in substantial morbidity and mortality. Critical intervention for infectious disease relies on rapid, sensitive and specific detection of the offending pathogen, and is central to the mission of microbiology laboratories at medical centers. Unfortunately, despite the recognition that outcomes from infectious illnesses are directly associated with time to pathogen recognition, as well as accurate identification of the class and species of microbe, and ability to identify the presence of drug resistance isolates, conventional hospital laboratories often remain encumbered by traditional slow multi-step culture based assays. Other limitations of the conventional laboratory which have become increasingly apparent include: extremely prolonged wait-times for pathogens with long generation time (up to several weeks); requirements for additional testing and wait times for speciation and identification of antimicrobial resistance; diminished test sensitivity for patients who have received antibiotics; and absolute inability to culture certain pathogens in disease states associated with microbial infection.

For more than a decade, molecular testing has been heralded as the diagnostic tool for the new millennium, whose ultimate potential could include forced obsolescence of traditional hospital laboratories. However, despite the fact that significant advances in clinical application of PCR techniques have occurred, the practicing physician still relies principally on standard techniques. A brief discussion of several existing applications of PCR in the hospital-based setting follows.

Generally speaking molecular diagnostics have been championed for identifying organisms that cannot be grown *in vitro*, or in instances where existing culture techniques are insensitive and/or require prolonged incubation times. PCR-based diagnostics have been successfully developed for a wide variety of microbes. Application to the clinical arena has met with variable success, with only a few assays achieving acceptance and utility.

One of the earliest, and perhaps most widely recognized applications of PCR for clinical practice is in detection of *Mycobacterium tuberculosis*. Clinical characteristics favoring development of a nonculture-based test for tuberculosis include week to month long delays associated with standard testing, occurrence of drug-resistant isolates and public health imperatives associated with recognition, isolation and treatment. Although frequently used as a diagnostic adjunctive, practical and routine clinical application of PCR remains problematic due to significant inter-laboratory variation in sensitivity, and inadequate specificity for use in low prevalence populations, requiring further development at the technical level. Recent advances in the laboratory suggest that identification of drug resistant isolates by amplification of mutations associated with specific antibiotic resistance (e.g., *rpoB* gene in rifampin resistant strains) may be forthcoming for clinical use, although widespread application will require extensive clinical validation.

One diagnostic assay, which has gained widespread acceptance, is for *C. trachomatis*. Conventional detection systems are limiting due to inadequate sensitivity and specificity (direct immunofluorescence or enzyme immunoassay) or the requirement for specialized culture facilities, due to the fastidious characteristics of this microbe. Laboratory development, followed by widespread clinical validation testing in a variety of acute and nonacute care settings have demonstrated excellent sensitivity (90-100%) and specificity (97%) of the PCR assay leading to its commercial development. Proven efficacy of the PCR assay from both genital and urine sampling, have resulted in its application to a variety of clinical setting, most recently including routine screening of patients considered at risk.

While the full potential for PCR diagnostics to provide rapid and critical information to physicians faced with difficult clinical-decisions has yet to be realized, one recently developed assay provides an example of the promise of this evolving technology. Distinguishing life-threatening causes of fever from more benign causes in children is a fundamental clinical dilemma faced by clinicians, particularly when infections of the central nervous system are being considered. Bacterial causes of meningitis can be highly aggressive, but generally cannot be differentiated on a clinical basis from aseptic meningitis, which is a relatively benign condition that can be managed on an outpatient basis. Existing blood culture methods often take several days to turn positive, and are often confounded by poor sensitivity or false-negative findings in patients receiving empiric antimicrobials. Testing and application of a PCR assay for enteroviral meningitis has been found to be highly sensitive. With reporting of results within 1 day, preliminary clinical trials have shown

significant reductions in hospital costs, due to decreased duration of hospital stays and reduction in antibiotic therapy. Other viral PCR assays, now routinely available include those for herpes simplex virus, cytomegalovirus, hepatitis and HIV. Each has a demonstrated cost savings role in clinical practice, including detection of otherwise difficult to diagnose infections and newly realized capacity to monitor progression of disease and response to therapy, vital in the management of chronic infectious diseases.

The concept of a universal detection system has been forwarded for identification of bacterial pathogens, and speaks most directly to the possible clinical implications of a broad-based screening tool for clinical use. Exploiting the existence of highly conserved regions of DNA common to all bacterial species in a PCR assay would empower physicians to rapidly identify the presence of bacteremia, which would profoundly impact patient care. Previous empiric decision making could be abandoned in favor of educated practice, allowing appropriate and expeditious decision-making regarding need for antibiotic therapy and hospitalization.

Experimental work using the conserved features of the 16S rRNA common to almost all bacterial species, is an area of active investigation. Hospital test sites have focused on "high yield" clinical settings where expeditious identification of the presence of systemic bacterial infection has immediate high morbidity and mortality consequences. Notable clinical infections have included evaluation of febrile infants at risk for sepsis, detection of bacteremia in febrile neutropenic cancer patients, and examination of critically ill patients in the intensive care unit. While several of these studies have reported promising results (with sensitivity and specificity well over 90%), significant technical difficulties (described below) remain, and have prevented general acceptance of this assay in clinics and hospitals (which remain dependent on standard blood culture methodologies). Even the revolutionary advances of real-time PCR technique, which offers a quantitative more reproducible and technically simpler system, remains encumbered by inherent technical limitations of the PCR assay.

The principle shortcomings of applying PCR assays to the clinical setting include: inability to eliminate background DNA contamination; interference with the PCR amplification by substrates present in the reaction; and limited capacity to provide rapid reliable speciation, antibiotic resistance and subtype identification. Some laboratories have recently made progress in identifying and removing inhibitors; however background contamination remains problematic, and methods directed towards eliminating exogenous

sources of DNA report significant diminution in assay sensitivity. Finally, while product identification and detailed characterization has been achieved using sequencing techniques, these approaches are laborious and time-intensive thus detracting from its clinical applicability.

Rapid and definitive microbial identification is desirable for a variety of industrial, medical, environmental, quality, and research reasons. Traditionally, the microbiology laboratory has functioned to identify the etiologic agents of infectious diseases through direct examination and culture of specimens. Since the mid-1980s, researchers have repeatedly demonstrated the practical utility of molecular biology techniques, many of which form the basis of clinical diagnostic assays. Some of these techniques include nucleic acid hybridization analysis, restriction enzyme analysis, genetic sequence analysis, and separation and purification of nucleic acids (See, e.g., J. Sambrook, E. F. Fritsch, and T. Maniatis, Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989). These procedures, in general, are time-consuming and tedious. Another option is the polymerase chain reaction (PCR) or other amplification procedure that amplifies a specific target DNA sequence based on the flanking primers used. Finally, detection and data analysis convert the hybridization event into an analytical result.

Other not yet fully realized applications of PCR for clinical medicine is the identification of infectious causes of disease previously described as idiopathic (e.g. *Bartonella henselae* in bacillary angiomatosis, and *Tropheryma whippellii* as the uncultured bacillus associated with Whipple's disease). Further, recent epidemiological studies which suggest a strong association between Chlamydia pneumonia and coronary artery disease, serve as example of the possible widespread, yet undiscovered links between pathogen and host which may ultimately allow for new insights into pathogenesis and novel life sustaining or saving therapeutics.

For the practicing clinician, PCR technology offers a yet unrealized potential for diagnostic omnipotence in the arena of infectious disease. A universal reliable infectious disease detection system would certainly become a fundamental tool in the evolving diagnostic armamentarium of the 21st century clinician. For front line emergency physicians, or physicians working in disaster settings, a quick universal detection system, would allow for molecular triage and early aggressive targeted therapy. Preliminary clinical studies using species specific probes suggest that implementing rapid testing in acute care setting is feasible. Resources could thus be appropriately applied, and patients with suspected

infections could rapidly be risk stratified to the different treatment settings, depending on the pathogen and virulence. Furthermore, links with data management systems, locally regionally and nationally, would allow for effective epidemiological surveillance, with obvious benefits for antibiotic selection and control of disease outbreaks.

For the hospitalists, the ability to speciate and subtype would allow for more precise decision-making regarding antimicrobial agents. Patients who are colonized with highly contagious pathogens could be appropriately isolated on entry into the medical setting without delay. Targeted therapy will diminish development of antibiotic resistance. Furthermore, identification of the genetic basis of antibiotic resistant strains would permit precise pharmacologic intervention. Both physician and patient would benefit with less need for repetitive testing and elimination of wait times for test results.

It is certain that the individual patient will benefit directly from this approach. Patients with unrecognized or difficult to diagnose infections would be identified and treated promptly. There will be reduced need for prolonged inpatient stays, with resultant decreases in iatrogenic events.

Mass spectrometry provides detailed information about the molecules being analyzed, including high mass accuracy. It is also a process that can be easily automated. Low-resolution MS may be unreliable when used to detect some known agents, if their spectral lines are sufficiently weak or sufficiently close to those from other living organisms in the sample. DNA chips with specific probes can only determine the presence or absence of specifically anticipated organisms. Because there are hundreds of thousands of species of benign bacteria, some very similar in sequence to threat organisms, even arrays with 10,000 probes lack the breadth needed to detect a particular organism.

Antibodies face more severe diversity limitations than arrays. If antibodies are designed against highly conserved targets to increase diversity, the false alarm problem will dominate, again because threat organisms are very similar to benign ones. Antibodies are only capable of detecting known agents in relatively uncluttered environments.

Several groups have described detection of PCR products using high resolution electrospray ionization-Fourier transform-ion cyclotron resonance mass spectrometry (ESI-FT-ICR MS). Accurate measurement of exact mass combined with knowledge of the number of at least one nucleotide allowed calculation of the total base composition for PCR duplex products of approximately 100 base pairs. (Aaserud *et al.*, *J. Am. Soc. Mass Spec.*, 1996, *7*, 1266-1269; Muddiman *et al.*, *Anal. Chem.*, 1997, *69*, 1543-1549; Wunschel *et al.*, *Anal.*

Chem., 1998, *70*, 1203-1207; Muddiman *et al.*, *Rev. Anal. Chem.*, 1998, *17*, 1-68). Electrospray ionization-Fourier transform-ion cyclotron resistance (ESI-FT-ICR) MS may be used to determine the mass of double-stranded, 500 base-pair PCR products via the average molecular mass (Hurst *et al.*, *Rapid Commun. Mass Spec.* 1996, *10*, 377-382). The use of matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry for characterization of PCR products has been described. (Muddiman *et al.*, *Rapid Commun. Mass Spec.*, 1999, *13*, 1201-1204). However, the degradation of DNAs over about 75 nucleotides observed with MALDI limited the utility of this method.

U.S. Patent No. 5,849,492 describes a method for retrieval of phylogenetically informative DNA sequences which comprise searching for a highly divergent segment of genomic DNA surrounded by two highly conserved segments, designing the universal primers for PCR amplification of the highly divergent region, amplifying the genomic DNA by PCR technique using universal primers, and then sequencing the gene to determine the identity of the organism.

U.S. Patent No. 5,965,363 discloses methods for screening nucleic acids for polymorphisms by analyzing amplified target nucleic acids using mass spectrometric techniques and to procedures for improving mass resolution and mass accuracy of these methods.

WO 99/14375 describes methods, PCR primers and kits for use in analyzing preselected DNA tandem nucleotide repeat alleles by mass spectrometry.

WO 98/12355 discloses methods of determining the mass of a target nucleic acid by mass spectrometric analysis, by cleaving the target nucleic acid to reduce its length, making the target single-stranded and using MS to determine the mass of the single-stranded shortened target. Also disclosed are methods of preparing a double-stranded target nucleic acid for MS analysis comprising amplification of the target nucleic acid, binding one of the strands to a solid support, releasing the second strand and then releasing the first strand which is then analyzed by MS. Kits for target nucleic acid preparation are also provided.

PCT WO97/33000 discloses methods for detecting mutations in a target nucleic acid by nonrandomly fragmenting the target into a set of single-stranded nonrandom length fragments and determining their masses by MS.

U.S. Patent No. 5,605,798 describes a fast and highly accurate mass spectrometer-based process for detecting the presence of a particular nucleic acid in a biological sample for diagnostic purposes.

WO 98/21066 describes processes for determining the sequence of a particular target nucleic acid by mass spectrometry. Processes for detecting a target nucleic acid present in a biological sample by PCR amplification and mass spectrometry detection are disclosed, as are methods for detecting a target nucleic acid in a sample by amplifying the target with primers that contain restriction sites and tags, extending and cleaving the amplified nucleic acid, and detecting the presence of extended product, wherein the presence of a DNA fragment of a mass different from wild-type is indicative of a mutation. Methods of sequencing a nucleic acid via mass spectrometry methods are also described.

WO 97/37041, WO 99/31278 and U.S. Patent No. 5,547,835 describe methods of sequencing nucleic acids using mass spectrometry. U.S. Patent Nos. 5,622,824, 5,872,003 and 5,691,141 describe methods, systems and kits for exonuclease-mediated mass spectrometric sequencing.

Thus, there is a need for a method for bioagent detection and identification which is both specific and rapid, and in which no nucleic acid sequencing is required. The present invention addresses this need.

SUMMARY OF THE INVENTION

The present invention is directed towards methods of identifying a pathogen in a biological sample by obtaining nucleic acid from a biological sample, selecting at least one pair of intelligent primers with the capability of amplification of nucleic acid of the pathogen, amplifying the nucleic acid with the primers to obtain at least one amplification product, determining the molecular mass of at least one amplification product from which the pathogen is identified. Further, this invention is directed to methods of epidemic surveillance. By identifying a pathogen from samples acquired from a plurality of geographic locations, the spread of the pathogen to a given geographic location can be determined.

The present invention is also directed to methods of diagnosis of a plurality of etiologic agents of disease in an individual by obtaining a biological sample from an individual, isolating nucleic acid from the biological sample, selecting a plurality of amplification primers with the capability of amplification of nucleic acid of a plurality of etiologic agents of disease, amplifying the nucleic acid with a plurality of primers to obtain a plurality of amplification products corresponding to a plurality of etiologic agents, determining the molecular masses of the plurality of unique amplification products which identify the members of the plurality of etiologic agents.

The present invention is also directed to methods of *in silico* screening of primer sets to be used in identification of a plurality of bioagents by preparing a base composition probability cloud plot from a plurality of base composition signatures of the plurality of bioagents generated *in silico*, inspecting the base composition probability cloud plot for overlap of clouds from different bioagents, and choosing primer sets based on minimal overlap of the clouds.

The present invention is also directed to methods of predicting the identity of a bioagent with a heretofore unknown base composition signature by preparing a base composition probability cloud plot from a plurality of base composition signatures of the plurality of bioagents which includes the heretofore unknown base composition, inspecting the base composition probability cloud for overlap of the heretofore unknown base composition with the cloud of a known bioagent such that overlap predicts that the identity of the bioagent with a heretofore unknown base composition signature equals the identity of the known bioagent.

The present invention is also directed to methods for determining a subspecies characteristic for a given pathogen in a biological sample by identifying the pathogen in a biological sample using broad range survey primers or division-wide primers, selecting at least one pair of drill-down primers to amplify nucleic acid segments which provide a subspecies characteristic about the pathogen, amplifying the nucleic acid segments to produce at least one drill-down amplification product and determining the base composition signature of the drill-down amplification product wherein the base composition signature provides a subspecies characteristic about the pathogen.

The present invention is also directed to methods of pharmacogenetic analysis by obtaining a sample of genomic DNA from an individual, selecting a segment of the genomic DNA which provides pharmacogenetic information, using at least one pair of intelligent primers to produce an amplification product which comprises the segment of genomic DNA and determining the base composition signature of the amplification product, wherein the base composition signature provides pharmacogenetic information about said individual.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-1H and Figure 2 are consensus diagrams that show examples of conserved regions from 16S rRNA (Fig. 1A-1, 1A-2, 1A-3, 1A-4, and 1A-5), 23S rRNA (3'-half, Fig. 1B, 1C, and 1D; 5'-half, Fig. 1E-F), 23S rRNA Domain I (Fig. 1G), 23S rRNA

Domain IV (Fig. 1H) and 16S rRNA Domain III (Fig. 2) which are suitable for use in the present invention. Lines with arrows are examples of regions to which intelligent primer pairs for PCR are designed. The label for each primer pair represents the starting and ending base number of the amplified region on the consensus diagram. Bases in capital letters are greater than 95% conserved; bases in lower case letters are 90-95% conserved, filled circles are 80-90% conserved; and open circles are less than 80% conserved. The label for each primer pair represents the starting and ending base number of the amplified region on the consensus diagram. The nucleotide sequence of the 16S rRNA consensus sequence is SEQ ID NO:3 and the nucleotide sequence of the 23S rRNA consensus sequence is SEQ ID NO:4.

Figure 2 shows a typical primer amplified region from the 16S rRNA Domain III shown in Figure 1A-1.

Figure 3 is a schematic diagram showing conserved regions in RNase P. Bases in capital letters are greater than 90% conserved; bases in lower case letters are 80-90% conserved; filled circles designate bases which are 70-80% conserved; and open circles designate bases that are less than 70% conserved.

Figure 4 is a schematic diagram of base composition signature determination using nucleotide analog "tags" to determine base composition signatures.

Figure 5 shows the deconvoluted mass spectra of a *Bacillus anthracis* region with and without the mass tag phosphorothioate A (A*). The two spectra differ in that the measured molecular weight of the mass tag-containing sequence is greater than the unmodified sequence.

Figure 6 shows base composition signature (BCS) spectra from PCR products from *Staphylococcus aureus* (*S. aureus* 16S_1337F) and *Bacillus anthracis* (*B. anthr.* 16S_1337F), amplified using the same primers. The two strands differ by only two (AT-->CG) substitutions and are clearly distinguished on the basis of their BCS.

Figure 7 shows that a single difference between two sequences (A14 in *B. anthracis* vs. A15 in *B. cereus*) can be easily detected using ESI-TOF mass spectrometry.

Figure 8 is an ESI-TOF of *Bacillus anthracis* spore coat protein sspE 56mer plus calibrant. The signals unambiguously identify *B. anthracis* versus other Bacillus species.

Figure 9 is an ESI-TOF of a *B. anthracis* synthetic 16S_1228 duplex (reverse and forward strands). The technique easily distinguishes between the forward and reverse strands.

Figure 10 is an ESI-FTICR-MS of a synthetic *B. anthracis* 16S_1337 46 base pair duplex.

Figure 11 is an ESI-TOF-MS of a 56mer oligonucleotide (3 scans) from the *B. anthracis* saspB gene with an internal mass standard. The internal mass standards are designated by asterisks.

Figure 12 is an ESI-TOF-MS of an internal standard with 5 mM TBA-TFA buffer showing that charge stripping with tributylammonium trifluoroacetate reduces the most abundant charge state from [M-8H+]8- to [M-3H+]3-.

Figure 13 is a portion of a secondary structure defining database according to one embodiment of the present invention, where two examples of selected sequences are displayed graphically thereunder.

Figure 14 is a three dimensional graph demonstrating the grouping of sample molecular weight according to species.

Figure 15 is a three dimensional graph demonstrating the grouping of sample molecular weights according to species of virus and mammal infected.

Figure 16 is a three dimensional graph demonstrating the grouping of sample molecular weights according to species of virus, and animal-origin of infectious agent.

Figure 17 is a figure depicting how the triangulation method of the present invention provides for the identification of an unknown bioagent without prior knowledge of the unknown agent. The use of different primer sets to distinguish and identify the unknown is also depicted as primer sets I, II and III within this figure. A three dimensional graph depicts all of bioagent space (**170**), including the unknown bioagent, which after use of primer set I (**171**) according to a method according to the present invention further differentiates and classifies bioagents according to major classifications (**176**) which, upon further analysis using primer set II (**172**) differentiates the unknown agent (**177**) from other, known agents (**173**) and finally, the use of a third primer set (**175**) further specifies subgroups within the family of the unknown (**174**).

Figure 18 shows a representative base composition probability cloud for a region of the RNA polymerase B gene from a cluster of enterobacteria. The dark spheres represent the actual base composition of the organisms. The lighter spheres represent the transitions among base compositions observed in different isolates of the same species of organism.

Figure 19 shows resolution of enterobacteriae members with primers targeting RNA polymerase B (rpoB). A single pair of primers targeting a hyper-variable region within rpoB was sufficient to resolve most members of this group at the genus level *(Salmonella from Escherichia* from *Yersinia*) as well as the species/strain level (*E. coli* K12 from O157). All

organisms with the exception of Y. *pestis* were tested in the lab and the measured base counts (shown with arrow) matched the predictions in every case.

Figure 20 shows detection of *S. aureus* in blood. Spectra on the right indicate signals corresponding to *S. aureus* detection in spiked wells A1 and A4 with no detection in control wells A2 and A3.

Figure 21 shows a representative base composition distribution of human adenovirus strain types for a single primer pair region on the hexon gene. The circles represent different adenovirus sequences in our database that were used for primer design. Measurement of masses and base counts for each of the unknown samples A, B, C and D matched one or more of the known groups of adenoviruses.

Figure 22 shows a representative broad range survey/drill-down process as applied to emm-typing of *streptococcus pyogenes* (Group A *Streptococcus*: GAS). Genetic material is extracted (**201**) and amplified using broad range survey primers (**202**). The amplification products are analyzed (**203**) to determine the presence and identity of bioagents at the species level. If *Streptococcus pyogenes* is detected (**204**), the emm-typing "drill-down" primers are used to reexamine the extract to identify the emm-type of the sample (**205**). Different sets of drill down primers can be employed to determine a subspecies characteristic for various strains of various bioagents (**206**).

Figure 23 shows a representative base composition distribution of bioagents detected in throat swabs from military personnel using a broad range primer pair directed to 16S rRNA.

Figure 24 shows a representative deconvoluted ESI-FTICR spectra of the PCR products produced by the gtr primer for samples 12 (top) and 10 (bottom) corresponding to emm types 3 and 6, respectively. Accurate mass measurements were obtained by using an internal mass standard and post-calibrating each spectrum; the experimental mass measurement uncertainty on each strand is + 0.035 Daltons (1 ppm). Unambiguous base compositions of the amplicons were determined by calculating all putative base compositions of each stand within the measured mass (and measured mass uncertainty) and selecting complementary pairs within the mass measurement uncertainty. In all cases there was only one base composition within 25 ppm. The measured mass difference of 15.985 Da between the strands shown on the left is in excellent agreement with the theoretical mass difference of 15.994 Da expected for an A to G substitution.

Figure 25 shows representative results of the base composition analysis on throat swab samples using the six primer pairs, 5'-emm gene sequencing and the MLST gene sequencing method of the present invention for an outbreak of *Streptococcus pyogenes* (group A *streptococcus*; GAS) at a military training camp.

Figure 26 shows: a) a representative ESI-FTICR mass spectrum of a restriction digest of a 986 bp region of the 16S ribosomal gene from E. coli K12 digested with a mixture of *BstNI, BsmFI, BfaI,* and *NcoI*; b) a deconvoluted representation (neutral mass) of the above spectrum showing the base compositions derived from accurate mass measurements of each fragment; and c) a representative reconstructed restriction map showing complete base composition coverage for nucleotides 1-856. The *Nco1* did not cut .

Figure 27 shows a representative base composition distribution of *poxviruses* for a single primer pair region on the DNA-dependent polymerase B gene (DdDpB). The spheres represent different *poxvirus* sequences that were used for primer design.

DESCRIPTION OF EMBODIMENTS

A. Introduction

The present invention provides, *inter alia*, methods for detection and identification of bioagents in an unbiased manner using "bioagent identifying amplicons." "Intelligent primers" are selected to hybridize to conserved sequence regions of nucleic acids derived from a bioagent and which bracket variable sequence regions to yield a bioagent identifying amplicon which can be amplified and which is amenable to molecular mass determination. The molecular mass then provides a means to uniquely identify the bioagent without a requirement for prior knowledge of the possible identity of the bioagent. The molecular mass or corresponding "base composition signature" (BCS) of the amplification product is then matched against a database of molecular masses or base composition signatures. Furthermore, the method can be applied to rapid parallel "multiplex" analyses, the results of which can be employed in a triangulation identification strategy. The present method provides rapid throughput and does not require nucleic acid sequencing of the amplified target sequence for bioagent detection and identification.

B. Bioagents

In the context of this invention, a "bioagent" is any organism, cell, or virus, living or dead, or a nucleic acid derived from such an organism, cell or virus. Examples of bioagents

include, but are not limited, to cells, including but not limited to, cells, including but not limited to human clinical samples, bacterial cells and other pathogens) viruses, fungi, and protists, parasites, and pathogenicity markers (including but not limited to: pathogenicity islands, antibiotic resistance genes, virulence factors, toxin genes and other bioregulating compounds). Samples may be alive or dead or in a vegetative state (for example, vegetative bacteria or spores) and may be encapsulated or bioengineered. In the context of this invention, a "pathogen" is a bioagent which causes a disease or disorder.

Despite enormous biological diversity, all forms of life on earth share sets of essential, common features in their genomes. Bacteria, for example have highly conserved sequences in a variety of locations on their genomes. Most notable is the universally conserved region of the ribosome. but there are also conserved elements in other non-coding RNAs, including RNAse P and the signal recognition particle (SRP) among others. Bacteria have a common set of absolutely required genes. About 250 genes are present in all bacterial species (*Proc. Natl. Acad. Sci. U.S.A.*, **1996**, *93*, 10268; *Science*, **1995**, *270*, 397), including tiny genomes like *Mycoplasma*, *Ureaplasma* and *Rickettsia*. These genes encode proteins involved in translation, replication, recombination and repair, transcription, nucleotide metabolism, amino acid metabolism, lipid metabolism, energy generation, uptake, secretion and the like. Examples of these proteins are DNA polymerase III beta, elongation factor TU, heat shock protein groEL, RNA polymerase beta, phosphoglycerate kinase, NADH dehydrogenase, DNA ligase, DNA topoisomerase and elongation factor G. Operons can also be targeted using the present method. One example of an operon is the bfp operon from enteropathogenic *E. coli*. Multiple core chromosomal genes can be used to classify bacteria at a genus or genus species level to determine if an organism has threat potential. The methods can also be used to detect pathogenicity markers (plasmid or chromosomal) and antibiotic resistance genes to confirm the threat potential of an organism and to direct countermeasures.

C. Selection of "Bioagent Identifying Amplicons"

Since genetic data provide the underlying basis for identification of bioagents by the methods of the present invention, it is necessary to select segments of nucleic acids which ideally provide enough variability to distinguish each individual bioagent and whose molecular mass is amenable to molecular mass determination. In one embodiment of the present invention, at least one polynucleotide segment is amplified to facilitate detection and

analysis in the process of identifying the bioagent. Thus, the nucleic acid segments which provide enough variability to distinguish each individual bioagent and whose molecular masses are amenable to molecular mass determination are herein described as "bioagent identifying amplicons." The term "amplicon" as used herein, refers to a segment of a polynucleotide which is amplified in an amplification reaction.

As used herein, "intelligent primers" are primers that are designed to bind to highly conserved sequence regions that flank an intervening variable region and yield amplification products which ideally provide enough variability to distinguish each individual bioagent, and which are amenable to molecular mass analysis. By the term "highly conserved," it is meant that the sequence regions exhibit between about 80-100%, or between about 90-100%, or between about 95-100% identity. The molecular mass of a given amplification product provides a means of identifying the bioagent from which it was obtained, due to the variability of the variable region. Thus design of intelligent primers requires selection of a variable region with appropriate variability to resolve the identity of a given bioagent.

In one embodiment, the bioagent identifying amplicon is a portion of a ribosomal RNA (rRNA) gene sequence. With the complete sequences of many of the smallest microbial genomes now available, it is possible to identify a set of genes that defines "minimal life" and identify composition signatures that uniquely identify each gene and organism. Genes that encode core life functions such as DNA replication, transcription, ribosome structure, translation, and transport are distributed broadly in the bacterial genome and are suitable regions for selection of bioagent identifying amplicons. Ribosomal RNA (rRNA) genes comprise regions that provide useful base composition signatures. Like many genes involved in core life functions, rRNA genes contain sequences that are extraordinarily conserved across bacterial domains interspersed with regions of high variability that are more specific to each species. The variable regions can be utilized to build a database of base composition signatures. The strategy involves creating a structure-based alignment of sequences of the small (16S) and the large (23S) subunits of the rRNA genes. For example, there are currently over 13,000 sequences in the ribosomal RNA database that has been created and maintained by Robin Gutell, University of Texas at Austin, and is publicly available on the Institute for Cellular and Molecular Biology web page on the world wide web of the Internet at, for example, "rna.icmb.utexas.edu/." There is also a publicly available rRNA database created and maintained by the University of Antwerp, Belgium on the world wide web of the Internet at, for example, "rrna.uia.ac.be."

These databases have been analyzed to determine regions that are useful as bioagent identifying amplicons. The characteristics of such regions include: a) between about 80 and 100%, or greater than about 95% identity among species of the particular bioagent of interest, of upstream and downstream nucleotide sequences which serve as sequence amplification primer sites; b) an intervening variable region which exhibits no greater than about 5% identity among species; and c) a separation of between about 30 and 1000 nucleotides, or no more than about 50-250 nucleotides, or no more than about 60-100 nucleotides, between the conserved regions.

As a non-limiting example, for identification of *Bacillus* species, the conserved sequence regions of the chosen bioagent identifying amplicon must be highly conserved among all *Bacillus* species while the variable region of the bioagent identifying amplicon is sufficiently variable such that the molecular masses of the amplification products of all species of *Bacillus* are distinguishable.

Bioagent identifying amplicons amenable to molecular mass determination are either of a length, size or mass compatible with the particular mode of molecular mass determination or compatible with a means of providing a predictable fragmentation pattern in order to obtain predictable fragments of a length compatible with the particular mode of molecular mass determination. Such means of providing a predictable fragmentation pattern of an amplification product include, but are not limited to, cleavage with restriction enzymes or cleavage primers, for example.

Identification of bioagents can be accomplished at different levels using intelligent primers suited to resolution of each individual level of identification. "Broad range survey" intelligent primers are designed with the objective of identifying a bioagent as a member of a particular division of bioagents. A "bioagent division" is defined as group of bioagents above the species level and includes but is not limited to: orders, families, classes, clades, genera or other such groupings of bioagents above the species level. As a non-limiting example, members of the *Bacillus/Clostridia* group or gamma-proteobacteria group may be identified as such by employing broad range survey intelligent primers such as primers which target 16S or 23S ribosomal RNA.

In some embodiments, broad range survey intelligent primers are capable of identification of bioagents at the species level. One main advantage of the detection methods of the present invention is that the broad range survey intelligent primers need not be specific for a particular bacterial species, or even genus, such as *Bacillus* or *Streptomyces*. Instead, the

primers recognize highly conserved regions across hundreds of bacterial species including, but not limited to, the species described herein. Thus, the same broad range survey intelligent primer pair can be used to identify any desired bacterium because it will bind to the conserved regions that flank a variable region specific to a single species, or common to several bacterial species, allowing unbiased nucleic acid amplification of the intervening sequence and determination of its molecular weight and base composition. For example, the 16S_971-1062, 16S_1228-1310 and 16S_1100-1188 regions are 98-99% conserved in about 900 species of bacteria (16S=16S rRNA, numbers indicate nucleotide position). In one embodiment of the present invention, primers used in the present method bind to one or more of these regions or portions thereof.

Due to their overall conservation, the flanking rRNA primer sequences serve as good intelligent primer binding sites to amplify the nucleic acid region of interest for most, if not all, bacterial species. The intervening region between the sets of primers varies in length and/or composition, and thus provides a unique base composition signature. Examples of intelligent primers that amplify regions of the 16S and 23S rRNA are shown in Figures 1A-1H. A typical primer amplified region in 16S rRNA is shown in Figure 2. The arrows represent primers that bind to highly conserved regions which flank a variable region in 16S rRNA domain III. The amplified region is the stem-loop structure under "1100-1188." It is advantageous to design the broad range survey intelligent primers to minimize the number of primers required for the analysis, and to allow detection of multiple members of a bioagent division using a single pair of primers. The advantage of using broad range survey intelligent primers is that once a bioagent is broadly identified, the process of further identification at species and sub-species levels is facilitated by directing the choice of additional intelligent primers.

"Division-wide" intelligent primers are designed with an objective of identifying a bioagent at the species level. As a non-limiting example, a *Bacillus anthracis, Bacillus cereus* and *Bacillus thuringiensis* can be distinguished from each other using division-wide intelligent primers. Division-wide intelligent primers are not always required for identification at the species level because broad range survey intelligent primers may provide sufficient identification resolution to accomplishing this identification objective.

"Drill-down" intelligent primers are designed with an objective of identifying a sub-species characteristic of a bioagent. A "sub-species characteristic" is defined as a property imparted to a bioagent at the sub-species level of identification as a result of the presence or

absence of a particular segment of nucleic acid. Such sub-species characteristics include, but are not limited to, strains, sub-types, pathogenicity markers such as antibiotic resistance genes, pathogenicity islands, toxin genes and virulence factors. Identification of such sub-species characteristics is often critical for determining proper clinical treatment of pathogen infections.

Chemical Modifications of Intelligent Primers

Ideally, intelligent primer hybridization sites are highly conserved in order to facilitate the hybridization of the primer. In cases where primer hybridization is less efficient due to lower levels of conservation of sequence, intelligent primers can be chemically modified to improve the efficiency of hybridization.

For example, because any variation (due to codon wobble in the 3rd position) in these conserved regions among species is likely to occur in the third position of a DNA triplet, oligonucleotide primers can be designed such that the nucleotide corresponding to this position is a base which can bind to more than one nucleotide, referred to herein as a "universal base." For example, under this "wobble" pairing, inosine (I) binds to U, C or A; guanine (G) binds to U or C, and uridine (U) binds to U or C. Other examples of universal bases include nitroindoles such as 5-nitroindole or 3-nitropyrrole (Loakes *et al.*, *Nucleosides and Nucleotides*, **1995**, *14*, 1001-1003), the degenerate nucleotides d**P** or d**K** (Hill *et al.*), an acyclic nucleoside analog containing 5-nitroindazole (Van Aerschot *et al.*, *Nucleosides and Nucleotides*, **1995**, *14*, 1053-1056) or the purine analog 1-(2-deoxy-β-D-ribofuranosyl)-imidazole-4-carboxamide (Sala *et al.*, *Nucl. Acids Res.*, **1996**, *24*, 3302-3306).

In another embodiment of the invention, to compensate for the somewhat weaker binding by the "wobble" base, the oligonucleotide primers are designed such that the first and second positions of each triplet are occupied by nucleotide analogs which bind with greater affinity than the unmodified nucleotide. Examples of these analogs include, but are not limited to, 2,6-diaminopurine which binds to thymine, propyne T which binds to adenine and propyne C and phenoxazines, including G-clamp, which binds to G. Propynylated pyrimidines are described in U.S. Patent Nos. 5,645,985, 5,830,653 and 5,484,908, each of which is commonly owned and incorporated herein by reference in its entirety. Propynylated primers are claimed in U.S Serial No. 10/294,203 which is also commonly owned and incorporated herein by reference in entirety. Phenoxazines are described in U.S. Patent Nos. 5,502,177, 5,763,588, and 6,005,096, each of which is incorporated herein by reference in its

entirety. G-clamps are described in U.S. Patent Nos. 6,007,992 and 6,028,183, each of which is incorporated herein by reference in its entirety.

D. Characterization of Bioagent Identifying Amplicons

A theoretically ideal bioagent detector would identify, quantify, and report the complete nucleic acid sequence of every bioagent that reached the sensor. The complete sequence of the nucleic acid component of a pathogen would provide all relevant information about the threat, including its identity and the presence of drug-resistance or pathogenicity markers. This ideal has not yet been achieved. However, the present invention provides a straightforward strategy for obtaining information with the same practical value based on analysis of bioagent identifying amplicons by molecular mass determination.

In some cases, a molecular mass of a given bioagent identifying amplicon alone does not provide enough resolution to unambiguously identify a given bioagent. For example, the molecular mass of the bioagent identifying amplicon obtained using the intelligent primer pair "16S_971" would be 55622 Da for both *E. coli* and *Salmonella typhimurium*. However, if additional intelligent primers are employed to analyze additional bioagent identifying amplicons, a "triangulation identification" process is enabled. For example, the "16S_1100" intelligent primer pair yields molecular masses of 55009 and 55005 Da for *E. coli* and *Salmonella typhimurium,* respectively. Furthermore, the "23S_855" intelligent primer pair yields molecular masses of 42656 and 42698 Da for *E. coli* and *Salmonella typhimurium,* respectively. In this basic example, the second and third intelligent primer pairs provided the additional "fingerprinting" capability or resolution to distinguish between the two bioagents.

In another embodiment, the triangulation identification process is pursued by measuring signals from a plurality of bioagent identifying amplicons selected within multiple core genes. This process is used to reduce false negative and false positive signals, and enable reconstruction of the origin of hybrid or otherwise engineered bioagents. In this process, after identification of multiple core genes, alignments are created from nucleic acid sequence databases. The alignments are then analyzed for regions of conservation and variation, and bioagent identifying amplicons are selected to distinguish bioagents based on specific genomic differences. For example, identification of the three part toxin genes typical of *B. anthracis* (Bowen *et al.*, *J. Appl. Microbiol.*, **1999,** *87*, 270-278) in the absence of the expected signatures from the *B. anthracis* genome would suggest a genetic engineering event.

The triangulation identification process can be pursued by characterization of bioagent identifying amplicons in a massively parallel fashion using the polymerase chain reaction (PCR), such as multiplex PCR, and mass spectrometric (MS) methods. Sufficient quantities of nucleic acids should be present for detection of bioagents by MS. A wide variety of techniques for preparing large amounts of purified nucleic acids or fragments thereof are well known to those of skill in the art. PCR requires one or more pairs of oligonucleotide primers that bind to regions which flank the target sequence(s) to be amplified. These primers prime synthesis of a different strand of DNA, with synthesis occurring in the direction of one primer towards the other primer. The primers, DNA to be amplified, a thermostable DNA polymerase (e.g. *Taq* polymerase), the four deoxynucleotide triphosphates, and a buffer are combined to initiate DNA synthesis. The solution is denatured by heating, then cooled to allow annealing of newly added primer, followed by another round of DNA synthesis. This process is typically repeated for about 30 cycles, resulting in amplification of the target sequence.

Although the use of PCR is suitable, other nucleic acid amplification techniques may also be used, including ligase chain reaction (LCR) and strand displacement amplification (SDA). The high-resolution MS technique allows separation of bioagent spectral lines from background spectral lines in highly cluttered environments.

In another embodiment, the detection scheme for the PCR products generated from the bioagent(s) incorporates at least three features. First, the technique simultaneously detects and differentiates multiple (generally about 6-10) PCR products. Second, the technique provides a molecular mass that uniquely identifies the bioagent from the possible primer sites. Finally, the detection technique is rapid, allowing multiple PCR reactions to be run in parallel.

E. Mass Spectrometric Characterization of Bioagent Identifying Amplicons

Mass spectrometry (MS)-based detection of PCR products provides a means for determination of BCS which has several advantages. MS is intrinsically a parallel detection scheme without the need for radioactive or fluorescent labels, since every amplification product is identified by its molecular mass. The current state of the art in mass spectrometry is such that less than femtomole quantities of material can be readily analyzed to afford information about the molecular contents of the sample. An accurate assessment of the molecular mass of the material can be quickly obtained, irrespective of whether the molecular

weight of the sample is several hundred, or in excess of one hundred thousand atomic mass units (amu) or Daltons. Intact molecular ions can be generated from amplification products using one of a variety of ionization techniques to convert the sample to gas phase. These ionization methods include, but are not limited to, electrospray ionization (ES), matrix-assisted laser desorption ionization (MALDI) and fast atom bombardment (FAB). For example, MALDI of nucleic acids, along with examples of matrices for use in MALDI of nucleic acids, are described in WO 98/54751 (Genetrace, Inc.).

In some embodiments, large DNAs and RNAs, or large amplification products therefrom, can be digested with restriction endonucleases prior to ionization. Thus, for example, an amplification product that was 10 kDa could be digested with a series of restriction endonucleases to produce a panel of, for example, 100 Da fragments. Restriction endonucleases and their sites of action are well known to the skilled artisan. In this manner, mass spectrometry can be performed for the purposes of restriction mapping.

Upon ionization, several peaks are observed from one sample due to the formation of ions with different charges. Averaging the multiple readings of molecular mass obtained from a single mass spectrum affords an estimate of molecular mass of the bioagent. Electrospray ionization mass spectrometry (ESI-MS) is particularly useful for very high molecular weight polymers such as proteins and nucleic acids having molecular weights greater than 10 kDa, since it yields a distribution of multiply-charged molecules of the sample without causing a significant amount of fragmentation.

The mass detectors used in the methods of the present invention include, but are not limited to, Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS), ion trap, quadrupole, magnetic sector, time of flight (TOF), Q-TOF, and triple quadrupole.

In general, the mass spectrometric techniques which can be used in the present invention include, but are not limited to, tandem mass spectrometry, infrared multiphoton dissociation and pyrolytic gas chromatography mass spectrometry (PGC-MS). In one embodiment of the invention, the bioagent detection system operates continually in bioagent detection mode using pyrolytic GC-MS without PCR for rapid detection of increases in biomass (for example, increases in fecal contamination of drinking water or of germ warfare agents). To achieve minimal latency, a continuous sample stream flows directly into the PGC-MS combustion chamber. When an increase in biomass is detected, a PCR process is automatically initiated. Bioagent presence produces elevated levels of large molecular fragments from, for example, about 100-7,000 Da which are observed in the PGC-MS

spectrum. The observed mass spectrum is compared to a threshold level and when levels of biomass are determined to exceed a predetermined threshold, the bioagent classification process described hereinabove (combining PCR and MS, such as FT-ICR MS) is initiated. Optionally, alarms or other processes (halting ventilation flow, physical isolation) are also initiated by this detected biomass level.

The accurate measurement of molecular mass for large DNAs is limited by the adduction of cations from the PCR reaction to each strand, resolution of the isotopic peaks from natural abundance ^{13}C and ^{15}N isotopes, and assignment of the charge state for any ion. The cations are removed by in-line dialysis using a flow-through chip that brings the solution containing the PCR products into contact with a solution containing ammonium acetate in the presence of an electric field gradient orthogonal to the flow. The latter two problems are addressed by operating with a resolving power of >100,000 and by incorporating isotopically depleted nucleotide triphosphates into the DNA. The resolving power of the instrument is also a consideration. At a resolving power of 10,000, the modeled signal from the $[M-14H+]^{14-}$ charge state of an 84mer PCR product is poorly characterized and assignment of the charge state or exact mass is impossible. At a resolving power of 33,000, the peaks from the individual isotopic components are visible. At a resolving power of 100,000, the isotopic peaks are resolved to the baseline and assignment of the charge state for the ion is straightforward. The $[^{13}C,^{15}N]$-depleted triphosphates are obtained, for example, by growing microorganisms on depleted media and harvesting the nucleotides (Batey *et al.*, *Nucl. Acids Res.*, **1992**, *20*, 4515-4523).

While mass measurements of intact nucleic acid regions are believed to be adequate to determine most bioagents, tandem mass spectrometry (MS^n) techniques may provide more definitive information pertaining to molecular identity or sequence. Tandem MS involves the coupled use of two or more stages of mass analysis where both the separation and detection steps are based on mass spectrometry. The first stage is used to select an ion or component of a sample from which further structural information is to be obtained. The selected ion is then fragmented using, e.g., blackbody irradiation, infrared multiphoton dissociation, or collisional activation. For example, ions generated by electrospray ionization (ESI) can be fragmented using IR multiphoton dissociation. This activation leads to dissociation of glycosidic bonds and the phosphate backbone, producing two series of fragment ions, called the *w*-series (having an intact 3' terminus and a 5' phosphate following internal cleavage) and the *a*-Base series(having an intact 5' terminus and a 3' furan).

The second stage of mass analysis is then used to detect and measure the mass of these resulting fragments of product ions. Such ion selection followed by fragmentation routines can be performed multiple times so as to essentially completely dissect the molecular sequence of a sample.

If there are two or more targets of similar molecular mass, or if a single amplification reaction results in a product which has the same mass as two or more bioagent reference standards, they can be distinguished by using mass-modifying "tags." In this embodiment of the invention, a nucleotide analog or "tag" is incorporated during amplification (e.g., a 5-(trifluoromethyl) deoxythymidine triphosphate) which has a different molecular weight than the unmodified base so as to improve distinction of masses. Such tags are described in, for example, PCT WO97/33000, which is incorporated herein by reference in its entirety. This further limits the number of possible base compositions consistent with any mass. For example, 5-(trifluoromethyl)deoxythymidine triphosphate can be used in place of dTTP in a separate nucleic acid amplification reaction. Measurement of the mass shift between a conventional amplification product and the tagged product is used to quantitate the number of thymidine nucleotides in each of the single strands. Because the strands are complementary, the number of adenosine nucleotides in each strand is also determined.

In another amplification reaction, the number of G and C residues in each strand is determined using, for example, the cytidine analog 5-methylcytosine (5-meC) or propyne C. The combination of the A/T reaction and G/C reaction, followed by molecular weight determination, provides a unique base composition. This method is summarized in Figure 4 and Table 1.

Table 1

Mass tag	Double strand sequence	Single strand Sequence	Total mass this strand	Base info this strand	Base info other strand	Total base comp. Top strand	Total base comp. Bottom strand
T*mass (T*-T) = x	T*ACGT*ACGT* AT*GCAT*GCA	T*ACGT*ACGT*	3x	3T	3A	3T 2A 2C 2G	3A 2T 2G 2C
		AT*GCAT*GCA	2x	2T	2A		

C*mass (C*-C) = y	TAC*GTAC*GT ATGC*ATGC*A	TAC*GTAC*GT	2x	2C	2G		
		ATGC*ATGC*A	2x	2C	2G		

The mass tag phosphorothioate A (A*) was used to distinguish a *Bacillus anthracis* cluster. The *B. anthracis* ($A_{14}G_9C_{14}T_9$) had an average MW of 14072.26, and the *B. anthracis* ($A_1A^*_{13}G_9C_{14}T_9$) had an average molecular weight of 14281.11 and the phosphorothioate A had an average molecular weight of +16.06 as determined by ESI-TOF MS. The deconvoluted spectra are shown in Figure 5.

In another example, assume the measured molecular masses of each strand are 30,000.115Da and 31,000.115 Da respectively, and the measured number of dT and dA residues are (30,28) and (28,30). If the molecular mass is accurate to 100 ppm, there are 7 possible combinations of dG+dC possible for each strand. However, if the measured molecular mass is accurate to 10 ppm, there are only 2 combinations of dG+dC, and at 1 ppm accuracy there is only one possible base composition for each strand.

Signals from the mass spectrometer may be input to a maximum-likelihood detection and classification algorithm such as is widely used in radar signal processing. The detection processing uses matched filtering of BCS observed in mass-basecount space and allows for detection and subtraction of signatures from known, harmless organisms, and for detection of unknown bioagent threats. Comparison of newly observed bioagents to known bioagents is also possible, for estimation of threat level, by comparing their BCS to those of known organisms and to known forms of pathogenicity enhancement, such as insertion of antibiotic resistance genes or toxin genes.

Processing may end with a Bayesian classifier using log likelihood ratios developed from the observed signals and average background levels. The program emphasizes performance predictions culminating in probability-of-detection versus probability-of-false-alarm plots for conditions involving complex backgrounds of naturally occurring organisms and environmental contaminants. Matched filters consist of a priori expectations of signal values given the set of primers used for each of the bioagents. A genomic sequence database (e.g. GenBank) is used to define the mass basecount matched filters. The database contains known threat agents and benign background organisms. The latter is used to estimate and subtract the signature produced by the background organisms. A maximum likelihood detection of known background organisms is implemented using matched filters and a

running-sum estimate of the noise covariance. Background signal strengths are estimated and used along with the matched filters to form signatures which are then subtracted. the maximum likelihood process is applied to this "cleaned up" data in a similar manner employing matched filters for the organisms and a running-sum estimate of the noise-covariance for the cleaned up data.

F. Base Composition Signatures as Indices of Bioagent Identifying Amplicons

Although the molecular mass of amplification products obtained using intelligent primers provides a means for identification of bioagents, conversion of molecular mass data to a base composition signature is useful for certain analyses. As used herein, a "base composition signature" (BCS) is the exact base composition determined from the molecular mass of a bioagent identifying amplicon. In one embodiment, a BCS provides an index of a specific gene in a specific organism.

Base compositions, like sequences, vary slightly from isolate to isolate within species. It is possible to manage this diversity by building "base composition probability clouds" around the composition constraints for each species. This permits identification of organisms in a fashion similar to sequence analysis. A "pseudo four-dimensional plot" can be used to visualize the concept of base composition probability clouds (Figure 18). Optimal primer design requires optimal choice of bioagent identifying amplicons and maximizes the separation between the base composition signatures of individual bioagents. Areas where clouds overlap indicate regions that may result in a misclassification, a problem which is overcome by selecting primers that provide information from different bioagent identifying amplicons, ideally maximizing the separation of base compositions. Thus, one aspect of the utility of an analysis of base composition probability clouds is that it provides a means for screening primer sets in order to avoid potential misclassifications of BCS and bioagent identity. Another aspect of the utility of base composition probability clouds is that they provide a means for predicting the identity of a bioagent whose exact measured BCS was not previously observed and/or indexed in a BCS database due to evolutionary transitions in its nucleic acid sequence.

It is important to note that, in contrast to probe-based techniques, mass spectrometry determination of base composition does not require prior knowledge of the composition in order to make the measurement, only to interpret the results. In this regard, the present invention provides bioagent classifying information similar to DNA sequencing and

phylogenetic analysis at a level sufficient to detect and identify a given bioagent. Furthermore, the process of determination of a previously unknown BCS for a given bioagent (for example, in a case where sequence information is unavailable) has downstream utility by providing additional bioagent indexing information with which to populate BCS databases. The process of future bioagent identification is thus greatly improved as more BCS indexes become available in the BCS databases.

Another embodiment of the present invention is a method of surveying bioagent samples that enables detection and identification of all bacteria for which sequence information is available using a set of twelve broad-range intelligent PCR primers. Six of the twelve primers are "broad range survey primers" herein defined as primers targeted to broad divisions of bacteria (for example, the *Bacillus/Clostridia* group or gamma-proteobacteria). The other six primers of the group of twelve primers are "division-wide" primers herein defined as primers which provide more focused coverage and higher resolution. This method enables identification of nearly 100% of known bacteria at the species level. A further example of this embodiment of the present invention is a method herein designated "survey/drill-down" wherein a subspecies characteristic for detected bioagents is obtained using additional primers. Examples of such a subspecies characteristic include but are not limited to: antibiotic resistance, pathogenicity island, virulence factor, strain type, sub-species type, and clade group. Using the survey/drill-down method, bioagent detection, confirmation and a subspecies characteristic can be provided within hours. Moreover, the survey/drill-down method can be focused to identify bioengineering events such as the insertion of a toxin gene into a bacterial species that does not normally make the toxin.

G. Fields of Application of the Present Invention

The present methods allow extremely rapid and accurate detection and identification of bioagents compared to existing methods. Furthermore, this rapid detection and identification is possible even when sample material is impure. The methods leverage ongoing biomedical research in virulence, pathogenicity, drug resistance and genome sequencing into a method which provides greatly improved sensitivity, specificity and reliability compared to existing methods, with lower rates of false positives. Thus, the methods are useful in a wide variety of fields, including, but not limited to, those fields discussed below.

1. Identification Of Pathogens In Humans And Animals

In other embodiments of the invention, the methods disclosed herein can identify infectious agents in biological samples. At least a first biological sample containing at least a first unidentified infectious agent is obtained. An identification analysis is carried out on the sample, whereby the first infectious agent in the first biological sample is identified. More particularly, a method of identifying an infectious agent in a biological entity is provided. An identification analysis is carried out on a first biological sample obtained from the biological entity, whereby at least one infectious agent in the biological sample from the biological entity is identified. The obtaining and the performing steps are, optionally, repeated on at least one additional biological sample from the biological entity.

The present invention also provides methods of identifying an infectious agent that is potentially the cause of a health condition in a biological entity. An identification analysis is carried out on a first test sample from a first infectious agent differentiating area of the biological entity, whereby at least one infectious agent is identified. The obtaining and the performing steps are, optionally, repeated on an additional infectious agent differentiating area of the biological entity.

Biological samples include, but are not limited to, hair, mucosa, skin, nail, blood, saliva, rectal, lung, stool, urine, breath, nasal, ocular sample, or the like. In some embodiments, one or more biological samples are analyzed by the methods described herein. The biological sample(s) contain at least a first unidentified infectious agent and may contain more than one infectious agent. The biological sample(s) are obtained from a biological entity. The biological sample can be obtained by a variety of manners such as by biopsy, swabbing, and the like. The biological samples may be obtained by a physician in a hospital or other health care environment. The physician may then perform the identification analysis or send the biological sample to a laboratory to carry out the analysis.

Biological entities include, but are not limited to, a mammal, a bird, or a reptile. The biological entity may be a cow, horse, dog, cat, or a primate. The biological entity can also be a human. The biological entity may be living or dead.

An infectious agent differentiating area is any area or location within a biological entity that can distinguish between a harmful versus normal health condition. An infectious agent differentiating area can be a region or area of the biological entity whereby an infectious agent is more likely to predominate from another region or area of the biological entity. For example, infectious agent differentiating areas may include the blood vessels of the heart (heart disease, coronary artery disease, etc.), particular portions of the digestive

system (ulcers, Crohn's disease, etc.), liver (hepatitis infections), and the like. In some embodiments, one or more biological samples from a plurality of infectious agent differentiating areas is analyzed the methods described herein.

Infectious agents of the invention may potentially cause a health condition in a biological entity. Health conditions include any condition, syndrome, illness, disease, or the like, identified currently or in the future by medical personnel. Infectious agents include, but are not limited to, bacteria, viruses, parasites, fungi, and the like.

In other embodiments of the invention, the methods disclosed herein can be used to screen blood and other bodily fluids and tissues for pathogenic and non-pathogenic bacteria, viruses, parasites, fungi and the like. Animal samples, including but not limited to, blood and other bodily fluid and tissue samples, can be obtained from living animals, who are either known or not known to or suspected of having a disease, infection, or condition. Alternately, animal samples such as blood and other bodily fluid and tissue samples can be obtained from deceased animals. Blood samples can be further separated into plasma or cellular fractions and further screened as desired. Bodily fluids and tissues can be obtained from any part of the animal or human body. Animal samples can be obtained from, for example, mammals and humans.

Clinical samples are analyzed for disease causing bioagents and biowarfare pathogens simultaneously with detection of bioagents at levels as low as 100-1000 genomic copies in complex backgrounds with throughput of approximately 100-300 samples with simultaneous detection of bacteria and viruses. Such analyses provide additional value in probing bioagent genomes for unanticipated modifications. These analyses are carried out in reference labs, hospitals and the LRN laboratories of the public health system in a coordinated fashion, with the ability to report the results via a computer network to a common data-monitoring center in real time. Clonal propagation of specific infectious agents, as occurs in the epidemic outbreak of infectious disease, can be tracked with base composition signatures, analogous to the pulse field gel electrophoresis fingerprinting patterns used in tracking the spread of specific food pathogens in the Pulse Net system of the CDC (Swaminathan, B., et al., *Emerging Infectious Diseases*, **2001**, *7*, 382-389). The present invention provides a digital barcode in the form of a series of base composition signatures, the combination of which is unique for each known organism. This capability enables real-time infectious disease monitoring across broad geographic locations, which may be essential in a simultaneous outbreak or attack in different cities.

In other embodiments of the invention, the methods disclosed herein can be used for detecting the presence of pathogenic and non-pathogenic bacteria, viruses, parasites, fungi and the like in organ donors and/or in organs from donors. Such examination can result in the prevention of the transfer of, for example, viruses such as West Nile virus, hepatitis viruses, human immunodeficiency virus, and the like from a donor to a recipient via a transplanted organ. The methods disclosed herein can also be used for detection of host versus graft or graft versus host rejection issues related to organ donors by detecting the presence of particular antigens in either the graft or host known or suspected of causing such rejection. In particular, the bioagents in this regard are the antigens of the major histocompatibility complex, such as the HLA antigens. The present methods can also be used to detect and track emerging infectious diseases, such as West Nile virus infection, HIV-related diseases.

In other embodiments of the invention, the methods disclosed herein can be used for pharmacogenetic analysis and medical diagnosis including, but not limited to, cancer diagnosis based on mutations and polymorphisms, drug resistance and susceptibility testing, screening for and/or diagnosis of genetic diseases and conditions, and diagnosis of infectious diseases and conditions. In context of the present invention, pharmacogenetics is defined as the study of variability in drug response due to genetic factors. Pharmacogenetic investigations are often based on correlating patient outcome with variations in genes involved in the mode of action of a given drug. For example, receptor genes, or genes involved in metabolic pathways. The methods of the present invention provide a means to analyze the DNA of a patient to provide the basis for pharmacogenetic analysis.

The present method can also be used to detect single nucleotide polymorphisms (SNPs), or multiple nucleotide polymorphisms, rapidly and accurately. A SNP is defined as a single base pair site in the genome that is different from one individual to another. The difference can be expressed either as a deletion, an insertion or a substitution, and is frequently linked to a disease state. Because they occur every 100-1000 base pairs, SNPs are the most frequently bound type of genetic marker in the human genome.

For example, sickle cell anemia results from an A-T transition, which encodes a valine rather than a glutamic acid residue. Oligonucleotide primers may be designed such that they bind to sequences that flank a SNP site, followed by nucleotide amplification and mass determination of the amplified product. Because the molecular masses of the resulting product from an individual who does not have sickle cell anemia is different from that of the product from an individual who has the disease, the method can be used to distinguish the

two individuals. Thus, the method can be used to detect any known SNP in an individual and thus diagnose or determine increased susceptibility to a disease or condition.

In one embodiment, blood is drawn from an individual and peripheral blood mononuclear cells (PBMC) are isolated and simultaneously tested, such as in a high-throughput screening method, for one or more SNPs using appropriate primers based on the known sequences which flank the SNP region. The National Center for Biotechnology Information maintains a publicly available database of SNPs on the world wide web of the Internet at, for example, "ncbi.nlm.nih.gov/SNP/."

The method of the present invention can also be used for blood typing. The gene encoding A, B or O blood type can differ by four single nucleotide polymorphisms. If the gene contains the sequence CGTGGTGACCCTT (SEQ ID NO:5), antigen A results. If the gene contains the sequence CGTCGTCACCGCTA (SEQ ID NO:6) antigen B results. If the gene contains the sequence CGTGGT-ACCCCTT (SEQ ID NO:7), blood group O results ("-" indicates a deletion). These sequences can be distinguished by designing a single primer pair which flanks these regions, followed by amplification and mass determination.

The method of the present invention can also be used for detection and identification of blood-borne pathogens such as *Staphylococcus aureus* for example.

The method of the present invention can also be used for strain typing of respiratory pathogens in epidemic surveillance. Group A streptococci (GAS), or *Streptococcus pyogenes,* is one of the most consequential causes of respiratory infections because of prevalence and ability to cause disease with complications such as acute rheumatic fever and acute glomerulonephritis . GAS also causes infections of the skin (*impetigo*) and, in rare cases, invasive disease such as necrotizing fasciitis and toxic shock syndrome. Despite many decades of study, the underlying microbial ecology and natural selection that favors enhanced virulence and explosive GAS outbreaks is still poorly understood. The ability to detect GAS and multiple other pathogenic and non-pathogenic bacteria and viruses in patient samples would greatly facilitate our understanding of GAS epidemics. It is also essential to be able to follow the spread of virulent strains of GAS in populations and to distinguish virulent strains from less virulent or avirulent streptococci that colonize the nose and throat of asymptomatic individuals at a frequency ranging from 5-20% of the population (Bisno, A. L. (1995) in Principles and Practice of Infectious Diseases, eds. Mandell, G. L., Bennett, J. E. & Dolin, R. (Churchill Livingston, New York), Vol. 2, pp. 1786-1799). Molecular methods have been developed to type GAS based upon the sequence of the emm gene that encodes the M-protein

virulence factor (Beall, B., Facklam, R. & Thompson, T. (1996) *J. Clin. Micro. 34*, 953-958; Beall, B., *et al.* (1997) J. *Clin. Micro. 35*, 1231-1235; Facklam, R., *et al.* (1999) *Emerging Infectious Diseases 5*, 247-253). Using this molecular classification, over 150 different emm-types are defined and correlated with phenotypic properties of thousands of GAS isolates (www.cdc.gov/ncidod/biotech/ strep/strepindex.html) (Facklam, R., *et al.* (2002) *Clinical Infectious Diseases 34*, 28-38). Recently, a strategy known as Multi Locus Sequence Typing (MLST) was developed to follow the molecular Epidemiology of GAS (13). In MLST, internal fragments of seven housekeeping genes are amplified, sequenced, and compared to a database of previously studied isolates (www.test.mlst.net/).

The present invention enables an emm-typing process to be carried out directly from throat swabs for a large number of samples within 12 hours, allowing strain tracking of an ongoing epidemic, even if geographically dispersed, on a larger scale than ever before achievable.

In another embodiment, the present invention can be employed in the serotyping of viruses including, but not limited to, adenoviruses. Adenoviruses are DNA viruses that cause over 50% of febrile respiratory illnesses in military recruits. Human adenoviruses are divided into six major serogroups (A through F), each containing multiple strain types. Despite the prevalence of adenoviruses, there are no rapid methods for detecting and serotyping adenoviruses.

In another embodiment, the present invention can be employed in distinguishing between members of the *Orthopoxvirus* genus. Smallpox is caused by the *Variola* virus. Other members of the genus include *Vaccinia, Monkeypox, Camelpox*, and *Cowpox*. All are capable of infecting humans, thus, a method capable of identifying and distinguishing among members of the *Orthopox* genus is a worthwhile objective.

In another embodiment, the present invention can be employed in distinguishing between viral agents of viral hemorrhagic fevers (VHF). VHF agents include, but are not limited to, *Filoviridae* (Marburg virus and Ebola virus), *Arenaviridae* (Lassa, Junin, Machupo, Sabia, and Guanarito viruses), *Bunyaviridae* (Crimean-Congo hemorrhagic fever virus (CCHFV), Rift Valley fever virus, and Hanta viruses), and *Flaviviridae* (yellow fever virus and dengue virus). Infections by VHF viruses are associated with a wide spectrum of clinical manifestations such as diarrhea, myalgia, cough, headache, pneumonia, encephalopathy, and hepatitis. Filoviruses, arenaviruses, and CCHFV are of particular relevance because they can be transmitted from human to human, thus causing epidemics

with high mortality rates (Khan, A.S., *et al.*, *Am. J. Trop. Med. Hyg.*, **1997**, *57*, 519-525). In the absence of bleeding or organ manifestation, VHF is clinically difficult to diagnose, and the various etiologic agents can hardly be distinguished by clinical tests. Current approaches to PCR detection of these agents are time-consuming, as they include a separate cDNA synthesis step prior to PCR, agarose gel analysis of PCR products, and in some instances a second round of nested amplification or Southern hybridization. PCRs for different pathogens have to be run assay by assay due to differences in cycling conditions, which complicate broad-range testing in a short period. Moreover, post-PCR processing or nested PCR steps included in currently used assays increase the risk of false positive results due to carryover contamination (Kwok, S. and R. Higuchi, *Nature* **1989**, *339*, 237-238).

In another embodiment, the present invention, can be employed in the diagnosis of a plurality of etiologic agents of a disease. An "etiologic agent" is herein defined as a pathogen acting as the causative agent of a disease. Diseases may be caused by a plurality of etiologic agents. For example, recent studies have implicated both human herpesvirus 6 (HHV-6) and the obligate intracellular bacterium *Chlamydia pneumoniae* in the etiology of multiple sclerosis (Swanborg, R.H. *Microbes and Infection* **2002**, *4*, 1327-1333). The present invention can be applied to the identification of multiple etiologic agents of a disease by, for example, the use of broad range bacterial intelligent primers and division-wide primers (if necessary) for the identification of bacteria such as *Chlamydia pneumoniae* followed by primers directed to viral housekeeping genes for the identification of viruses such as HHV-6, for example.

In other embodiments of the invention, the methods disclosed herein can be used for detection and identification of pathogens in livestock. Livestock includes, but is not limited to, cows, pigs, sheep, chickens, turkeys, goats, horses and other farm animals. For example, conditions classified by the California Department of Food and Agriculture as emergency conditions in livestock (www.cdfa.ca.gov/ahfss/ah/pdfs/CA_reportable_disease_list_05292002.pdf) include, but are not limited to: Anthrax (*Bacillus anthracis*), Screwworm myiasis (*Cochliomyia hominivorax* or *Chrysomya bezziana*), African trypanosomiasis (Tsetse fly diseases), Bovine babesiosis (*piroplasmosis*), Bovine spongiform encephalopathy (Mad Cow), Contagious bovine pleuropneumonia (*Mycoplasma mycoides mycoides* small colony), Foot-and-mouth disease (Hoof-and-mouth), Heartwater (*Cowdria ruminantium*), Hemorrhagic septicemia (*Pasteurella multocida* serotypes B:2 or E:2), Lumpy skin disease, Malignant catarrhal fever (African type), Rift Valley fever, Rinderpest (Cattle plague),

Theileriosis (Corridor disease, East Coast fever), Vesicular stomatitis, Contagious agalactia (*Mycoplasma* species), Contagious caprine pleuropneumonia (*Mycoplasma capricolum capripneumoniae*), Nairobi sheep disease, Peste des petits ruminants (Goat plague), Pulmonary adenomatosis (Viral neoplastic pneumonia), *Salmonella abortus ovis*, Sheep and goat pox, African swine fever, Classical swine fever (Hog cholera), Japanese encephalitis, Nipah virus, Swine vesicular disease, Teschen disease (*Enterovirus encephalomyelitis*), Vesicular exanthema, Exotic Newcastle disease (Viscerotropic velogenic Newcastle disease), Highly pathogenic avian influenza (Fowl plague), African horse sickness, Dourine (*Trypanosoma equiperdum*), Epizootic lymphangitis (equine blastomycosis, equine histoplasmosis), Equine piroplasmosis (*Babesia equi, B. caballi*), Glanders (Farcy) (*Pseudomonas mallei*), Hendra virus (Equine *morbillivirus*), Horse pox, Surra (Trypanosoma evansi), Venezuelan equine encephalomyelitis, West Nile Virus, Chronic wasting disease in cervids, and Viral hemorrhagic disease of rabbits (calicivirus)

Conditions classified by the California Department of Food and Agriculture as regulated conditions in livestock include, but are not limited to: rabies, Bovine brucellosis (*Brucella abortus*), Bovine tuberculosis (Mycobac*terium bovis*), Cattle scabies (multiple types), Trichomonosis (*Tritrichomonas fetus*), Caprine and ovine brucellosis (excluding Brucella ovis), Scrapie, Sheep scabies (Body mange) (*Psoroptes ovis*), Porcine brucellosis (*Brucella suis*), Pseudorabies (Aujeszky's disease), Ornithosis (*Psittacosis* or avian *chlamydiosis*) (*Chlamydia psittaci*), Pullorum disease (Fowl typhoid) (*Salmonella gallinarum* and *pullorum*), Contagious equine metritis (*Taylorella equigenitalis*), Equine encephalomyelitis (Eastern and Western equine encephalitis), Equine infectious anemia (Swamp fever), Duck viral enteritis (Duck plague), and Tuberculosis in cervids.

Additional conditions monitored by the California Department of Food and Agriculture include, but are not limited to: Avian tuberculosis (*Mycobacterium avium*), Echinococcosis/Hydatidosis (*Echinococcus* species), Leptospirosis, Anaplasmosis (*Anaplasma marginale* or *A. centrale*), Bluetongue, Bovine cysticercosis (*Taenia saginata* in humans), Bovine genital campylobacteriosis (*Campylobacter fetus venerealis*), Dermatophilosis (Streptothricosis, mycotic dermatitis) (*Dermatophilus congolensis*), Enzootic bovine leukosis (Bovine leukemia virus), Infectious bovine rhinotracheitis (Bovine herpesvirus-1), Johne's disease (Paratuberculosis) (*Mycobacterium avium paratuberculosis*), Malignant catarrhal fever (North American), Q Fever (Coxiella burnetii), Caprine (contagious) arthritis/encephalitis, Enzootic abortion of ewes (Ovine chlamydiosis)

(Chlamydia psittaci), Maedi-Visna (Ovine progressive pneumonia), Atrophic rhinitis (*Bordetella bronchiseptica, Pasteurella multocida*), Porcine cysticercosis (*Taenia solium* in humans), Porcine reproductive and respiratory syndrome, Transmissible gastroenteritis (*coronavirus*), Trichinellosis (*Trichinella spiralis*), Avian infectious bronchitis, Avian infectious laryngotracheitis, Duck viral hepatitis, Fowl cholera (*Pasteurella multocida*), Fowl pox, Infectious bursal disease (Gumboro disease), Low pathogenic avian influenza, Marek's disease, Mycoplasmosis (*Mycoplasma gallisepticum*), Equine influenza Equine rhinopneumonitis (Equine herpesvirus-1), Equine viral arteritis, and Horse mange (multiple types).

2. Identification of Bioagents of Biological Warfare

A key problem in determining that an infectious outbreak is the result of a bioterrorist attack is the sheer variety of organisms that might be used by terrorists. According to a recent review (Taylor, L. H. *et al. Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2001**, *356*, 983-989), there are over 1400 organisms infectious to humans; most of these have the potential to be used in a deliberate, malicious attack. These numbers do not include numerous strain variants of each organism, bioengineered versions, or pathogens that infect plants or animals. Paradoxically, most of the new technology being developed for detection of biological weapons incorporates a version of quantitative PCR, which is based upon the use of highly specific primers and probes designed to selectively identify specific pathogenic organisms. This approach requires assumptions about the type and strain of bacteria or virus which is expected to be detected. Although this approach will work for the most obvious organisms, like smallpox and anthrax, experience has shown that it is very difficult to anticipate what a terrorist will do.

The present invention can be used to detect and identify any biological agent, including bacteria, viruses, fungi and toxins without prior knowledge of the organism being detected and identified. As one example, where the agent is a biological threat, the information obtained such as the presence of toxin genes, pathogenicity islands and antibiotic resistance genes for example, is used to determine practical information needed for countermeasures. In addition, the methods can be used to identify natural or deliberate engineering events including chromosome fragment swapping, molecular breeding (gene shuffling) and emerging infectious diseases. The present invention provides broad-function technology that may be the only practical means for rapid diagnosis of disease caused by a

biowarfare or bioterrorist attack, especially an attack that might otherwise be missed or mistaken for a more common infection.

Bacterial biological warfare agents capable of being detected by the present methods include, but are not limited to, *Bacillus anthracis* (anthrax), *Yersinia pestis* (pneumonic plague), *Franciscella tularensis* (tularemia), *Brucella suis*, *Brucella abortus*, *Brucella melitensis* (undulant fever), *Burkholderia mallei* (glanders), *Burkholderia pseudomalleii* (melioidosis), *Salmonella typhi* (typhoid fever), *Rickettsia typhii* (epidemic typhus), *Rickettsia prowasekii* (endemic typhus) and *Coxiella burnetii* (Q fever), *Rhodobacter capsulatus, Chlamydia pneumoniae*, *Escherichia coli*, *Shigella dysenteriae*, *Shigella flexneri*, *Bacillus cereus*, *Clostridium botulinum*, *Coxiella burnetti*, *Pseudomonas aeruginosa*, *Legionella pneumophila*, and *Vibrio cholerae*.

Besides 16S and 23S rRNA, other target regions suitable for use in the present invention for detection of bacteria include, but are not limited to, 5S rRNA and RNase P (Figure 3).

Fungal biowarfare agents include, but are not limited to, *Coccidioides immitis* (Coccidioidomycosis), and *Magnaporthe grisea.*

Biological warfare toxin genes capable of being detected by the methods of the present invention include, but are not limited to, botulinum toxin, T-2 mycotoxins, ricin, staph enterotoxin B, shigatoxin, abrin, aflatoxin, *Clostridium perfringens* epsilon toxin, conotoxins, diacetoxyscirpenol, tetrodotoxin and saxitoxin.

Parasites that could be used in biological warfare include, but are not limited to: *Ascaris suum, Giardia lamblia, Cryptosporidium,* and *Schistosoma.*

Biological warfare viral threat agents are mostly RNA viruses (positive-strand and negative–strand), with the exception of smallpox. Every RNA virus is a family of related viruses (quasispecies). These viruses mutate rapidly and the potential for engineered strains (natural or deliberate) is very high. RNA viruses cluster into families that have conserved RNA structural domains on the viral genome (e.g., virion components, accessory proteins) and conserved housekeeping genes that encode core viral proteins including, for single strand positive strand RNA viruses, RNA-dependent RNA polymerase, double stranded RNA helicase, chymotrypsin-like and papain-like proteases and methyltransferases. "Housekeeping genes" refers to genes that are generally always expressed and thought to be involved in routine cellular metabolism.

Examples of (-)-strand RNA viruses include, but are not limited to, arenaviruses (e.g., sabia virus, lassa fever, Machupo, Argentine hemorrhagic fever, flexal virus), bunyaviruses (e.g., hantavirus, nairovirus, phlebovirus, hantaan virus, Congo-crimean hemorrhagic fever, rift valley fever), and mononegavirales (e.g., filovirus, paramyxovirus, ebola virus, Marburg, equine morbillivirus).

Examples of (+)-strand RNA viruses include, but are not limited to, picornaviruses (e.g., coxsackievirus, echovirus, human coxsackievirus A, human echovirus, human enterovirus, human poliovirus, hepatitis A virus, human parechovirus, human rhinovirus), astroviruses (e.g., human astrovirus), calciviruses (e.g., chiba virus, chitta virus, human calcivirus, norwalk virus), nidovirales (e.g., human coronavirus, human torovirus), flaviviruses (e.g., dengue virus 1-4, Japanese encephalitis virus, Kyanasur forest disease virus, Murray Valley encephalitis virus, Rocio virus, St. Louis encephalitis virus, West Nile virus, yellow fever virus, hepatitis c virus) and togaviruses (e.g., Chikugunya virus, Eastern equine encephalitis virus, Mayaro virus, O'nyong-nyong virus, Ross River virus, Venezuelan equine encephalitis virus, Rubella virus, hepatitis E virus). The hepatitis C virus has a 5'-untranslated region of 340 nucleotides, an open reading frame encoding 9 proteins having 3010 amino acids and a 3'-untranslated region of 240 nucleotides. The 5'-UTR and 3'-UTR are 99% conserved in hepatitis C viruses.

In one embodiment, the target gene is an RNA-dependent RNA polymerase or a helicase encoded by (+)-strand RNA viruses, or RNA polymerase from a (-)-strand RNA virus. (+)-strand RNA viruses are double stranded RNA and replicate by RNA-directed RNA synthesis using RNA-dependent RNA polymerase and the positive strand as a template. Helicase unwinds the RNA duplex to allow replication of the single stranded RNA. These viruses include viruses from the family picornaviridae (e.g., poliovirus, coxsackievirus, echovirus), togaviridae (e.g., alphavirus, flavivirus, rubivirus), arenaviridae (e.g., lymphocytic choriomeningitis virus, lassa fever virus), cononaviridae (e.g., human respiratory virus) and Hepatitis A virus. The genes encoding these proteins comprise variable and highly conserved regions which flank the variable regions.

In one embodiment, the method can be used to detect the presence of antibiotic resistance and/or toxin genes in a bacterial species. For example, *Bacillus anthracis* comprising a tetracycline resistance plasmid and plasmids encoding one or both anthracis toxins (px01 and/or px02) can be detected by using antibiotic resistance primer sets and toxin

gene primer sets. If the *B. anthracis* is positive for tetracycline resistance, then a different antibiotic, for example quinalone, is used.

While the present invention has been described with specificity in accordance with certain of its embodiments, the following examples serve only to illustrate the invention and are not intended to limit the same.

EXAMPLES

Example 1: Nucleic Acid Isolation and PCR

In one embodiment, nucleic acid is isolated from the organisms and amplified by PCR using standard methods prior to BCS determination by mass spectrometry. Nucleic acid is isolated, for example, by detergent lysis of bacterial cells, centrifugation and ethanol precipitation. Nucleic acid isolation methods are described in, for example, *Current Protocols in Molecular Biology* (Ausubel et al.) and *Molecular Cloning; A Laboratory Manual* (Sambrook *et al.*). The nucleic acid is then amplified using standard methodology, such as PCR, with primers which bind to conserved regions of the nucleic acid which contain an intervening variable sequence as described below.

General Genomic DNA Sample Prep Protocol: Raw samples are filtered using Supor-200 0.2 μm membrane syringe filters (VWR International) . Samples are transferred to 1.5 ml eppendorf tubes pre-filled with 0.45 g of 0.7 mm Zirconia beads followed by the addition of 350 μl of ATL buffer (Qiagen, Valencia, CA). The samples are subjected to bead beating for 10 minutes at a frequency of 19 l/s in a Retsch Vibration Mill (Retsch). After centrifugation, samples are transferred to an S-block plate (Qiagen) and DNA isolation is completed with a BioRobot 8000 nucleic acid isolation robot (Qiagen).

Swab Sample Protocol: Allegiance S/P brand culture swabs and collection/transport system are used to collect samples. After drying, swabs are placed in 17x100 mm culture tubes (VWR International) and the genomic nucleic acid isolation is carried out automatically with a Qiagen Mdx robot and the Qiagen QIAamp DNA Blood BioRobot Mdx genomic preparation kit (Qiagen, Valencia, CA).

Example 2: Mass spectrometry

FTICR Instrumentation: The FTICR instrument is based on a 7 tesla actively shielded superconducting magnet and modified Bruker Daltonics Apex II 70e ion optics and vacuum chamber. The spectrometer is interfaced to a LEAP PAL autosampler and a custom

fluidics control system for high throughput screening applications. Samples are analyzed directly from 96-well or 384-well microtiter plates at a rate of about 1 sample/minute. The Bruker data-acquisition platform is supplemented with a lab-built ancillary NT datastation which controls the autosampler and contains an arbitrary waveform generator capable of generating complex rf-excite waveforms (frequency sweeps, filtered noise, stored waveform inverse Fourier transform (SWIFT), etc.) for sophisticated tandem MS experiments. For oligonucleotides in the 20-30-mer regime typical performance characteristics include mass resolving power in excess of 100,000 (FWHM), low ppm mass measurement errors, and an operable *m/z* range between 50 and 5000 *m/z*.

Modified ESI Source: In sample-limited analyses, analyte solutions are delivered at 150 nL/minute to a 30 mm i.d. fused-silica ESI emitter mounted on a 3-D micromanipulator. The ESI ion optics consists of a heated metal capillary, an rf-only hexapole, a skimmer cone, and an auxiliary gate electrode. The 6.2 cm rf-only hexapole is comprised of 1 mm diameter rods and is operated at a voltage of 380 Vpp at a frequency of 5 MHz. A lab-built electro-mechanical shutter can be employed to prevent the electrospray plume from entering the inlet capillary unless triggered to the "open" position via a TTL pulse from the data station. When in the "closed" position, a stable electrospray plume is maintained between the ESI emitter and the face of the shutter. The back face of the shutter arm contains an elastomeric seal that can be positioned to form a vacuum seal with the inlet capillary. When the seal is removed, a 1 mm gap between the shutter blade and the capillary inlet allows constant pressure in the external ion reservoir regardless of whether the shutter is in the open or closed position. When the shutter is triggered, a "time slice" of ions is allowed to enter the inlet capillary and is subsequently accumulated in the external ion reservoir. The rapid response time of the ion shutter (< 25 ms) provides reproducible, user defined intervals during which ions can be injected into and accumulated in the external ion reservoir.

Apparatus for Infrared Multiphoton Dissociation: A 25 watt CW CO_2 laser operating at 10.6 μm has been interfaced to the spectrometer to enable infrared multiphoton dissociation (IRMPD) for oligonucleotide sequencing and other tandem MS applications. An aluminum optical bench is positioned approximately 1.5 m from the actively shielded superconducting magnet such that the laser beam is aligned with the central axis of the magnet. Using standard IR-compatible mirrors and kinematic mirror mounts, the unfocused 3 mm laser beam is aligned to traverse directly through the 3.5 mm holes in the trapping electrodes of the FTICR trapped ion cell and longitudinally traverse the hexapole region of

the external ion guide finally impinging on the skimmer cone. This scheme allows IRMPD to be conducted in an *m/z* selective manner in the trapped ion cell (e.g. following a SWIFT isolation of the species of interest), or in a broadband mode in the high pressure region of the external ion reservoir where collisions with neutral molecules stabilize IRMPD-generated metastable fragment ions resulting in increased fragment ion yield and sequence coverage.

Example 3 :Identification of Bioagents

Table 2 shows a small cross section of a database of calculated molecular masses for over 9 primer sets and approximately 30 organisms. The primer sets were derived from rRNA alignment. Examples of regions from rRNA consensus alignments are shown in Figures 1A-1C. Lines with arrows are examples of regions to which intelligent primer pairs for PCR are designed. The primer pairs are >95% conserved in the bacterial sequence database (currently over 10,000 organisms). The intervening regions are variable in length and/or composition, thus providing the base composition "signature" (BCS) for each organism. Primer pairs were chosen so the total length of the amplified region is less than about 80-90 nucleotides. The label for each primer pair represents the starting and ending base number of the amplified region on the consensus diagram.

Included in the short bacterial database cross-section in Table 2 are many well known pathogens/biowarfare agents (shown in bold/red typeface) such as *Bacillus anthracis* or *Yersinia pestis* as well as some of the bacterial organisms found commonly in the natural environment such as *Streptomyces*. Even closely related organisms can be distinguished from each other by the appropriate choice of primers. For instance, two low G+C organisms, *Bacillus anthracis* and *Staph aureus*, can be distinguished from each other by using the primer pair defined by 16S_1337 or 23S_855 (ΔM of 4 Da).

Table 2: Cross Section Of A Database Of Calculated Molecular Masses[1]

Primer Regions ---> / Bug Name	16S_971	16S_1100	16S_1337	16S_1294	16S_1228	23S_1021	23S_855	23S_193	23S_115
Acinetobacter calcoaceticus	55619.1	55004	28446.7	35854.9	51295.4	30299	42654	39557.5	54999
Bacillus anthracis	**55005**	**54388**	**28448**	**35238**	**51296**	**30295**	**42651**	**39560**	**56850**
Bacillus cereus	55622.1	54387.9	28447.6	35854.9	51296.4	30295	42651	39560.5	56850.3
Bordetella bronchiseptica	56857.3	51300.4	28446.7	35857.9	51307.4	30299	42653	39559.5	51920.5
Borrelia burgdorferi	56231.2	55621.1	28440.7	35852.9	51295.4	30297	42029.9	38941.4	52524.6
Brucella abortus	**58098**	**55011**	**28448**	**35854**	**50683**				
Campylobacter jejuni	58088.5	54386.9	29061.8	35856.9	50674.3	30294	42032.9	39558.5	45732.5
Chlamydia pnuemoniae	**55000**	**55007**	**29063**	**35855**	**50676**	**30295**	**42036**	**38941**	**56230**
Clostridium botulinum	**55006**	**53767**	**28445**	**35855**	**51291**	**30300**	**42656**	**39562**	**54999**
Clostridium difficile	56855.3	54386.9	28444.7	35853.9	51296.4	30294	41417.8	39556.5	55612.2
Enterococcus faecalis	55620.1	54387.9	28447.6	35858.9	51296.4	30297	42652	39559.5	56849.3
Escherichia coli	**55622**	**55009**	**28445**	**35857**	**51301**	**30301**	**42656**	**39562**	**54999**
Francisella tularensis	**53769**	**54385**	**28445**	**35856**	**51298**				
Haemophilus influenzae	55620.1	55006	28444.7	35855.9	51298.4	30298	42656	39560.5	55613.1
Klebsiella pneumoniae	55622.1	55008	28442.7	35856.9	51297.4	30300	42655	39562.5	55000
Legionella pneumophila	**55618**	**55626**	**28446**	**35857**	**51303**				
Mycobacterium avium	54390.9	55631.1	29064.8	35858.9	51915.5	30298	42656	38942.4	56241.2
Mycobacterium leprae	54389.9	55629.1	29064.8	35860.9	51917.5	30298	42656	39559.5	56240.2
Mycobacterium tuberculosis	54390.9	55629.1	29064.8	35860.9	51301.4	30299	42656	39560.5	56243.2
Mycoplasma genitalium	53143.7	45115.4	29061.8	35854.9	50671.3	30294	43264.1	39558.5	56842.4
Mycoplasma pneumoniae	53143.7	45118.4	29061.8	35854.9	50673.3	30294	43264.1	39559.5	56843.4
Neisseria gonorrhoeae	55627.1	54389.9	28445.7	35855.9	51302.4	30300	42649	39561.5	55000
Pseudomonas aeruginosa	**55623**	**55010**	**28443**	**35858**	**51301**	**30298**	**43272**	**39558**	**55619**
Rickettsia prowazekii	**58093**	**55621**	**28448**	**35853**	**50677**	**30293**	**42650**	**39559**	**53139**
Rickettsia rickettsii	**58094**	**55623**	**28448**	**35853**	**50679**	**30293**	**42648**	**39559**	**53755**
Salmonella typhimurium	**55622**	**55005**	**28445**	**35857**	**51301**	**30301**	**42658**		
Shigella dysenteriae	**55623**	**55009**	**28444**	**35857**	**51301**				
Staphylococcus aureus	56854.3	54386.9	28443.7	35852.9	51294.4	30298	42655	39559.5	57466.4
Streptomyces	54389.9	59341.6	29063.8	35858.9	51300.4			39563.5	56864.3
Treponema pallidum	56245.2	55631.1	28445.7	35851.9	51297.4	30299	42034.9	38939.4	57473.4
Vibrio cholerae	**55625**	**55626**	**28443**	**35857**	**52536**	**29063**	**30303**	**35241**	**50675**
Vibrio parahaemolyticus	54384.9	55626.1	28444.7	34620.7	50064.2				
Yersinia pestis	**55620**	**55626**	**28443**	**35857**	**51299**				

[1]Molecular mass distribution of PCR amplified regions for a selection of organisms (rows) across various primer pairs (columns). Pathogens are shown in **bold.** Empty cells indicate presently incomplete or missing data.

Figure 6 shows the use of ESI-FT-ICR MS for measurement of exact mass. The spectra from 46mer PCR products originating at position 1337 of the 16S rRNA from *S. aureus* (upper) and *B. anthracis* (lower) are shown. These data are from the region of the spectrum containing signals from the $[M-8H+]^{8-}$ charge states of the respective 5'-3' strands. The two strands differ by two (AT→CG) substitutions, and have measured masses of 14206.396 and 14208.373 + 0.010 Da, respectively. The possible base compositions derived from the masses of the forward and reverse strands for the *B. anthracis* products are listed in Table 3.

Table 3: Possible base composition for *B. anthracis* products

Calc. Mass	Error	Base Comp.
14208.2935	0.079520	A1 G17 C10 T18
14208.3160	0.056980	A1 G20 C15 T10

14208.3386	0.034440	A1 G23 C20 T2
14208.3074	0.065560	A6 G11 C3 T26
14208.3300	0.043020	A6 G14 C8 T18
14208.3525	0.020480	A6 G17 C13 T10
14208.3751	0.002060	A6 G20 C18 T2
14208.3439	0.029060	A11 G8 C1 T26
14208.3665	0.006520	A11 G11 C6 T18
14208.3890	**0.016020**	**A11 G14 C11 T10**
14208.4116	0.038560	A11 G17 C16 T2
14208.4030	0.029980	A16 G8 C4 T18
14208.4255	0.052520	A16 G11 C9 T10
14208.4481	0.075060	A16 G14 C14 T2
14208.4395	0.066480	A21 G5 C2 T18
14208.4620	0.089020	A21 G8 C7 T10
14079.2624	0.080600	A0 G14 C13 T19
14079.2849	0.058060	A0 G17 C18 T11
14079.3075	0.035520	A0 G20 C23 T3
14079.2538	0.089180	A5 G5 C1 T35
14079.2764	0.066640	A5 G8 C6 T27
14079.2989	0.044100	A5 G11 C11 T19
14079.3214	0.021560	A5 G14 C16 T11
14079.3440	0.000980	A5 G17 C21 T3
14079.3129	0.030140	A10 G5 C4 T27
14079.3354	0.007600	A10 G8 C9 T19
14079.3579	**0.014940**	**A10 G11 C14 T11**
14079.3805	0.037480	A10 G14 C19 T3
14079.3494	0.006360	A15 G2 C2 T27
14079.3719	0.028900	A15 G5 C7 T19
14079.3944	0.051440	A15 G8 C12 T11
14079.4170	0.073980	A15 G11 C17 T3
14079.4084	0.065400	A20 G2 C5 T19
14079.4309	0.087940	A20 G5 C10 T13

Among the 16 compositions for the forward strand and the 18 compositions for the reverse strand that were calculated, only one pair (shown in **bold**) are complementary, corresponding to the actual base compositions of the *B. anthracis* PCR products.

Example 4: BCS of Region from *Bacillus anthracis* and *Bacillus cereus*

A conserved Bacillus region from *B. anthracis* ($A_{14}G_9C_{14}T_9$) and *B. cereus* ($A_{15}G_9C_{13}T_9$) having a C to A base change was synthesized and subjected to ESI-TOF MS. The results are shown in Figure 7 in which the two regions are clearly distinguished using the method of the present invention (MW=14072.26 vs. 14096.29).

Example 5: Identification of additional bioagents

In other examples of the present invention, the pathogen *Vibrio cholera* can be distinguished from *Vibrio parahemolyticus* with $\Delta M > 600$ Da using one of three 16S primer sets shown in Table 2 (16S_971, 16S_1228 or 16S_1294) as shown in Table 4. The two mycoplasma species in the list (*M. genitalium* and *M. pneumoniae*) can also be distinguished from each other, as can the three mycobacteriae. While the direct mass measurements of amplified products can identify and distinguish a large number of organisms, measurement of the base composition signature provides dramatically enhanced resolving power for closely related organisms. In cases such as *Bacillus anthracis* and *Bacillus cereus* that are virtually indistinguishable from each other based solely on mass differences, compositional analysis or fragmentation patterns are used to resolve the differences. The single base difference between the two organisms yields different fragmentation patterns, and despite the presence of the ambiguous/unidentified base N at position 20 in *B. anthracis*, the two organisms can be identified.

Tables 4a-b show examples of primer pairs from Table 1 which distinguish pathogens from background.

Table 4a

Organism name	**23S_855**	**16S_1337**	**23S_1021**
Bacillus anthracis	42650.98	28447.65	30294.98
Staphylococcus aureus	42654.97	28443.67	30297.96

Table 4b

Organism name	16S_971	16S_1294	16S_1228
Vibrio cholerae	55625.09	35856.87	52535.59
Vibrio parahaemolyticus	54384.91	34620.67	50064.19

Table 5 shows the expected molecular weight and base composition of region 16S_1100-1188 in *Mycobacterium avium* and *Streptomyces sp.*

Table 5

Region	Organism name	Length	Molecular weight	Base comp.
16S_1100-1188	*Mycobacterium avium*	82	25624.1728	$A_{16}G_{32}C_{18}T_{16}$
16S_1100-1188	*Streptomyces sp.*	96	29904.871	$A_{17}G_{38}C_{27}T_{14}$

Table 6 shows base composition (single strand) results for 16S_1100-1188 primer amplification reactions different species of bacteria. Species which are repeated in the table (e.g., *Clostridium botulinum*) are different strains which have different base compositions in the 16S_1100-1188 region.

Table 6

Organism name	Base comp.	Organism name	Base comp.
Mycobacterium avium	$A_{16}G_{32}C_{18}T_{16}$	*Vibrio cholerae*	$A_{23}G_{30}C_{21}T_{16}$
Streptomyces sp.	$A_{17}G_{38}C_{27}T_{14}$	***Aeromonas hydrophila***	$\mathbf{A_{23}G_{31}C_{21}T_{15}}$
Ureaplasma urealyticum	$A_{18}G_{30}C_{17}T_{17}$	***Aeromonas salmonicida***	$\mathbf{A_{23}G_{31}C_{21}T_{15}}$
Streptomyces sp.	$A_{19}G_{36}C_{24}T_{18}$	*Mycoplasma genitalium*	$A_{24}G_{19}C_{12}T_{18}$
Mycobacterium leprae	$A_{20}G_{32}C_{22}T_{16}$	*Clostridium botulinum*	$A_{24}G_{25}C_{18}T_{20}$
M. tuberculosis	$\mathbf{A_{20}G_{33}C_{21}T_{16}}$	*Bordetella bronchiseptica*	$A_{24}G_{26}C_{19}T_{14}$
Nocardia asteroides	$\mathbf{A_{20}G_{33}C_{21}T_{16}}$	*Francisella tularensis*	$A_{24}G_{26}C_{19}T_{19}$
Fusobacterium necroforum	$A_{21}G_{26}C_{22}T_{18}$	***Bacillus anthracis***	$\mathbf{A_{24}G_{26}C_{20}T_{18}}$
Listeria monocytogenes	$A_{21}G_{27}C_{19}T_{19}$	***Campylobacter jejuni***	$\mathbf{A_{24}G_{26}C_{20}T_{18}}$
Clostridium botulinum	$A_{21}G_{27}C_{19}T_{21}$	***Staphylococcus aureus***	$\mathbf{A_{24}G_{26}C_{20}T_{18}}$
Neisseria gonorrhoeae	$A_{21}G_{28}C_{21}T_{18}$	*Helicobacter pylori*	$A_{24}G_{26}C_{20}T_{19}$
Bartonella quintana	$A_{21}G_{30}C_{22}T_{16}$	*Helicobacter pylori*	$A_{24}G_{26}C_{21}T_{18}$
Enterococcus faecalis	$A_{22}G_{27}C_{20}T_{19}$	*Moraxella catarrhalis*	$A_{24}G_{26}C_{23}T_{16}$

Bacillus megaterium	$A_{22}G_{28}C_{20}T_{18}$	*Haemophilus influenzae Rd*	$A_{24}G_{28}C_{20}T_{17}$
Bacillus subtilis	$A_{22}G_{28}C_{21}T_{17}$	***Chlamydia trachomatis***	$\mathbf{A_{24}G_{28}C_{21}T_{16}}$
Pseudomonas aeruginosa	$A_{22}G_{29}C_{23}T_{15}$	***Chlamydophila pneumoniae***	$\mathbf{A_{24}G_{28}C_{21}T_{16}}$
Legionella pneumophila	$A_{22}G_{32}C_{20}T_{16}$	***C. pneumonia AR39***	$\mathbf{A_{24}G_{28}C_{21}T_{16}}$
Mycoplasma pneumoniae	$A_{23}G_{20}C_{14}T_{16}$	*Pseudomonas putida*	$A_{24}G_{29}C_{21}T_{16}$
Clostridium botulinum	$A_{23}G_{26}C_{20}T_{19}$	***Proteus vulgaris***	$\mathbf{A_{24}G_{30}C_{21}T_{15}}$
Enterococcus faecium	$A_{23}G_{26}C_{21}T_{18}$	***Yersinia pestis***	$\mathbf{A_{24}G_{30}C_{21}T_{15}}$
Acinetobacter calcoaceti	$A_{23}G_{26}C_{21}T_{19}$	***Yersinia pseudotuberculos***	$\mathbf{A_{24}G_{30}C_{21}T_{15}}$
Leptospira borgpeterseni	$\mathbf{A_{23}G_{26}C_{24}T_{15}}$	*Clostridium botulinum*	$A_{25}G_{24}C_{18}T_{21}$
Leptospira interrogans	$\mathbf{A_{23}G_{26}C_{24}T_{15}}$	*Clostridium tetani*	$A_{25}G_{25}C_{18}T_{20}$
Clostridium perfringens	$A_{23}G_{27}C_{19}T_{19}$	*Francisella tularensis*	$A_{25}G_{25}C_{19}T_{19}$
Bacillus anthracis	$\mathbf{A_{23}G_{27}C_{20}T_{18}}$	*Acinetobacter calcoacetic*	$A_{25}G_{26}C_{20}T_{19}$
Bacillus cereus	$\mathbf{A_{23}G_{27}C_{20}T_{18}}$	*Bacteriodes fragilis*	$A_{25}G_{27}C_{16}T_{22}$
Bacillus thuringiensis	$\mathbf{A_{23}G_{27}C_{20}T_{18}}$	*Chlamydophila psittaci*	$A_{25}G_{27}C_{21}T_{16}$
Aeromonas hydrophila	$A_{23}G_{29}C_{21}T_{16}$	*Borrelia burgdorferi*	$A_{25}G_{29}C_{17}T_{19}$
Escherichia coli	$A_{23}G_{29}C_{21}T_{16}$	*Streptobacillus monilifor*	$A_{26}G_{26}C_{20}T_{16}$
Pseudomonas putida	$A_{23}G_{29}C_{21}T_{17}$	*Rickettsia prowazekii*	$A_{26}G_{28}C_{18}T_{18}$
Escherichia coli	$\mathbf{A_{23}G_{29}C_{22}T_{15}}$	*Rickettsia rickettsii*	$A_{26}G_{28}C_{20}T_{16}$
Shigella dysenteriae	$\mathbf{A_{23}G_{29}C_{22}T_{15}}$	*Mycoplasma mycoides*	$A_{28}G_{23}C_{16}T_{20}$

The same organism having different base compositions are different strains. Groups of organisms which are highlighted or in italics have the same base compositions in the amplified region. Some of these organisms can be distinguished using multiple primers. For example, *Bacillus anthracis* can be distinguished from *Bacillus cereus* and *Bacillus thuringiensis* using the primer 16S_971-1062 (Table 7). Other primer pairs which produce unique base composition signatures are shown in Table 6 (bold). Clusters containing very similar threat and ubiquitous non-threat organisms (e.g. *anthracis* cluster) are distinguished at high resolution with focused sets of primer pairs. The known biowarfare agents in Table 6 are *Bacillus anthracis, Yersinia pestis, Francisella tularensis* and *Rickettsia prowazekii.*

Table 7

Organism	16S_971-1062	16S_1228-1310	16S_1100-1188
Aeromonas hydrophila	$A_{21}G_{29}C_{22}T_{20}$	$A_{22}G_{27}C_{21}T_{13}$	$A_{23}G_{31}C_{21}T_{15}$
Aeromonas salmonicida	$A_{21}G_{29}C_{22}T_{20}$	$A_{22}G_{27}C_{21}T_{13}$	$A_{23}G_{31}C_{21}T_{15}$
Bacillus anthracis	$\mathbf{A_{21}G_{27}C_{22}T_{22}}$	$A_{24}G_{22}C_{19}T_{18}$	$A_{23}G_{27}C_{20}T_{18}$
Bacillus cereus	$A_{22}G_{27}C_{21}T_{22}$	$A_{24}G_{22}C_{19}T_{18}$	$A_{23}G_{27}C_{20}T_{18}$
Bacillus thuringiensis	$A_{22}G_{27}C_{21}T_{22}$	$A_{24}G_{22}C_{19}T_{18}$	$A_{23}G_{27}C_{20}T_{18}$
Chlamydia trachomatis	$\mathbf{A_{22}G_{26}C_{20}T_{23}}$	$\mathbf{A_{24}G_{23}C_{19}T_{16}}$	$A_{24}G_{28}C_{21}T_{16}$
Chlamydia pneumoniae AR39	$A_{26}G_{23}C_{20}T_{22}$	$A_{26}G_{22}C_{16}T_{18}$	$A_{24}G_{28}C_{21}T_{16}$
Leptospira borgpetersenii	$A_{22}G_{26}C_{20}T_{21}$	$A_{22}G_{25}C_{21}T_{15}$	$A_{23}G_{26}C_{24}T_{15}$
Leptospira interrogans	$A_{22}G_{26}C_{20}T_{21}$	$A_{22}G_{25}C_{21}T_{15}$	$A_{23}G_{26}C_{24}T_{15}$
Mycoplasma genitalium	$A_{28}G_{23}C_{15}T_{22}$	$\mathbf{A_{30}G_{18}C_{15}T_{19}}$	$\mathbf{A_{24}G_{19}C_{12}T_{18}}$
Mycoplasma pneumoniae	$A_{28}G_{23}C_{15}T_{22}$	$\mathbf{A_{27}G_{19}C_{16}T_{20}}$	$\mathbf{A_{23}G_{20}C_{14}T_{16}}$
Escherichia coli	$\mathbf{A_{22}G_{28}C_{20}T_{22}}$	$A_{24}G_{25}C_{21}T_{13}$	$A_{23}G_{29}C_{22}T_{15}$
Shigella dysenteriae	$\mathbf{A_{22}G_{28}C_{21}T_{21}}$	$A_{24}G_{25}C_{21}T_{13}$	$A_{23}G_{29}C_{22}T_{15}$
Proteus vulgaris	$\mathbf{A_{23}G_{26}C_{22}T_{21}}$	$\mathbf{A_{26}G_{24}C_{19}T_{14}}$	$A_{24}G_{30}C_{21}T_{15}$
Yersinia pestis	$A_{24}G_{25}C_{21}T_{22}$	$A_{25}G_{24}C_{20}T_{14}$	$A_{24}G_{30}C_{21}T_{15}$
Yersinia pseudotuberculosis	$A_{24}G_{25}C_{21}T_{22}$	$A_{25}G_{24}C_{20}T_{14}$	$A_{24}G_{30}C_{21}T_{15}$
Francisella tularensis	$\mathbf{A_{20}G_{25}C_{21}T_{23}}$	$\mathbf{A_{23}G_{26}C_{17}T_{17}}$	$\mathbf{A_{24}G_{26}C_{19}T_{19}}$
Rickettsia prowazekii	$\mathbf{A_{21}G_{26}C_{24}T_{25}}$	$\mathbf{A_{24}G_{23}C_{16}T_{19}}$	$\mathbf{A_{26}G_{28}C_{18}T_{18}}$
Rickettsia rickettsii	$\mathbf{A_{21}G_{26}C_{25}T_{24}}$	$\mathbf{A_{24}G_{24}C_{17}T_{17}}$	$\mathbf{A_{26}G_{28}C_{20}T_{16}}$

The sequence of *B. anthracis* and *B. cereus* in region 16S_971 is shown below. Shown in bold is the single base difference between the two species which can be detected using the methods of the present invention. *B. anthracis* has an ambiguous base at position 20.

*B.anthracis*_16S_971

GCGAAGAACCUUACCAGGU*N*UUGACAUCCUCUGA**C**AACCCUAGAGAUAGGGCU UCUCCUUCGGGAGCAGAGUGACAGGUGGUGCAUGGUU (SEQ ID NO:1)

*B.cereus*_16S_971
GCGAAGAACCUUACCAGGUCUUGACAUCCUCUGAAAACCCUAGAGAUAGGGCU
UCUCCUUCGGGAGCAGAGUGACAGGUGGUGCAUGGUU (SEQ ID NO:2)

Example 6: ESI-TOF MS of sspE 56-mer Plus Calibrant

The mass measurement accuracy that can be obtained using an internal mass standard in the ESI-MS study of PCR products is shown in Fig.8. The mass standard was a 20-mer phosphorothioate oligonucleotide added to a solution containing a 56-mer PCR product from the *B. anthracis* spore coat protein sspE. The mass of the expected PCR product distinguishes *B. anthracis* from other species of Bacillus such as *B. thuringiensis* and *B. cereus*.

Example 7: *B. anthracis* ESI-TOF Synthetic 16S_1228 Duplex

An ESI-TOF MS spectrum was obtained from an aqueous solution containing 5 μM each of synthetic analogs of the expected forward and reverse PCR products from the nucleotide 1228 region of the *B. anthracis* 16S rRNA gene. The results (Fig. 9) show that the molecular weights of the forward and reverse strands can be accurately determined and easily distinguish the two strands. The $[M-21H^+]^{21-}$ and $[M-20H^+]^{20-}$ charge states are shown.

Example 8: ESI-FTICR-MS of Synthetic *B. anthracis* 16S_1337 46 Base Pair Duplex

An ESI-FTICR-MS spectrum was obtained from an aqueous solution containing 5 μM each of synthetic analogs of the expected forward and reverse PCR products from the nucleotide 1337 region of the *B. anthracis* 16S rRNA gene. The results (Fig. 10) show that the molecular weights of the strands can be distinguished by this method. The $[M-16H^+]^{16-}$ through $[M-10H^+]^{10-}$ charge states are shown. The insert highlights the resolution that can be realized on the FTICR-MS instrument, which allows the charge state of the ion to be determined from the mass difference between peaks differing by a single 13C substitution.

Example 9: ESI-TOF MS of 56-mer Oligonucleotide from saspB Gene of *B. anthracis* with Internal Mass Standard

ESI-TOF MS spectra were obtained on a synthetic 56-mer oligonucleotide (5 μM) from the saspB gene of *B. anthracis* containing an internal mass standard at an ESI of 1.7 μL/min as a function of sample consumption. The results (Fig. 11) show that the signal to

noise is improved as more scans are summed, and that the standard and the product are visible after only 100 scans.

Example 10: ESI-TOF MS of an Internal Standard with Tributylammonium (TBA)-trifluoroacetate (TFA) Buffer

An ESI-TOF-MS spectrum of a 20-mer phosphorothioate mass standard was obtained following addition of 5 mM TBA-TFA buffer to the solution. This buffer strips charge from the oligonucleotide and shifts the most abundant charge state from $[M-8H^+]^{8-}$ to $[M-3H^+]^{3-}$ (Fig. 12).

Example 11: Master Database Comparison

The molecular masses obtained through Examples 1-10 are compared to molecular masses of known bioagents stored in a master database to obtain a high probability matching molecular mass.

Example 12: Master Data Base Interrogation over the Internet

The same procedure as in Example 11 is followed except that the local computer did not store the Master database. The Master database is interrogated over an internet connection, searching for a molecular mass match.

Example 13: Master Database Updating

The same procedure as in example 11 is followed except the local computer is connected to the internet and has the ability to store a master database locally. The local computer system periodically, or at the user's discretion, interrogates the Master database, synchronizing the local master database with the global Master database. This provides the current molecular mass information to both the local database as well as to the global Master database. This further provides more of a globalized knowledge base.

Example 14: Global Database Updating

The same procedure as in example 13 is followed except there are numerous such local stations throughout the world. The synchronization of each database adds to the diversity of information and diversity of the molecular masses of known bioagents.

Various modifications of the invention, in addition to those described herein, will be apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims. Each reference cited in the present application is incorporated herein by reference in its entirety.

Example 15: Demonstration of Detection and Identification of Five Species of Bacteria in a Mixture

Broad range intelligent primers were chosen following analysis of a large collection of curated bacterial 16S rRNA sequences representing greater than 4000 species of bacteria. Examples of primers capable of priming from greater than 90% of the organisms in the collection include, but are not limited to, those exhibited in Table 8 wherein Tp = 5'propynylated uridine and Cp = 5'propynylated cytidine.

Table 8: Intelligent Primer Pairs for Identification of Bacteria

Primer Pair Name	Forward Primer Sequence	Forward SEQ ID NO:	Reverse Primer Sequence	Reverse SEQ ID NO:
16S_EC_1077_1195	GTGAGATGTTGGGTTAAGTCCCGTAACGAG	8	GACGTCATCCCCACCTTCCTC	9
16S_EC_1082_1197	ATGTTGGGTTAAGTCCCGCAACGAG	10	TTGACGTCATCCCCACCTTCCTC	11
16S_EC_1090_1196	TTAAGTCCCGCAACGATCGCAA	12	TGACGTCATCCCCACCTTCCTC	13
16S_EC_1222_1323	GCTACACACGTGCTACAATG	14	CGAGTTGCAGACTGCGATCCG	15
16S_EC_1332_1407	AAGTCGGAATCGCTAGTAATCG	16	GACGGGCGGTGTGTACAAG	17
16S_EC_30_126	TGAACGCTGGTGGCATGCTTAACAC	18	TACGCATTACTCACCCGTCCGC	19
16S_EC_38_120	GTGGCATGCCTAATACATGCAAGTCG	20	TTACTCACCCGTCCGCCGCT	21
16S_EC_49_120	TAACACATGCAAGTCGAACG	22	TTACTCACCCGTCCGCC	23
16S_EC_683_795	GTGTAGCGGTGAAATGCG	24	GTATCTAATCCTGTTTGCTCCC	25
16S_EC_713_809	AGAACACCGATGGCGAAGGC	26	CGTGGACTACCAGGGTATCTA	27
16S_EC_785_897	GGATTAGAGACCCTGGTAGTCC	28	GGCCGTACTCCCCAGGCG	29
16S_EC_785_897_2	GGATTAGATACCCTGGTAGTCCACGC	30	GGCCGTACTCCCCAGGCG	31
16S_EC_789_894	TAGATACCCTGGTAGTCCACGC	32	CGTACTCCCCAGGCG	33
16S_EC_960_1073	TTCGATGCAACGCGAAGAACCT	34	ACGAGCTGACGACAGCCATG	35
16S_EC_969_1078	ACGCGAAGAACCTTACC	36	ACGACACGAGCTGACGAC	37

23S_EC_1826_1924	CTGACACCTGCCCGGTGC	38	GACCGTTATAGTTACGGCC	39
23S_EC_2645_2761	TCTGTCCCTAGTACGAGAGGACCGG	40	TGCTTAGATGCTTTCAGC	41
23S_EC_2645_2767	CTGTCCCTAGTACGAGAGGACCGG	42	GTTTCATGCTTAGATGCTTTCAGC	43
23S_EC_493_571	GGGGAGTGAAAGAGATCCTGAAACCG	44	ACAAAAGGTACGCCGTCACCC	45
23S_EC_493_571_2	GGGGAGTGAAAGAGATCCTGAAACCG	46	ACAAAAGGCACGCCATCACCC	47
23S_EC_971_1077	CGAGAGGGAAACAACCCAGACC	48	TGGCTGCTTCTAAGCCAAC	49
INFB_EC_1365_1467	TGCTCGTGGTGCACAAGTAACGGATATTA	50	TGCTGCTTTCGCATGGTTAATTGCTTCAA	51
RPOC_EC_1018_1124	CAAAACTTATTAGGTAAGCGTGTTGACT	52	TCAAGCGCCATTTCTTTTGGTAAACCACAT	53
RPOC_EC_1018_1124_2	CAAAACTTATTAGGTAAGCGTGTTGACT	54	TCAAGCGCCATCTCTTTCGGTAATCCACAT	55
RPOC_EC_114_232	TAAGAAGCCGGAAACCATCAACTACCG	56	GGCGCTTGTACTTACCGCAC	57
RPOC_EC_2178_2246	TGATTCTGGTGCCCGTGGT	58	TTGGCCATCAGGCCACGCATAC	59
RPOC_EC_2178_2246_2	TGATTCCGGTGCCCGTGGT	60	TTGGCCATCAGACCACGCATAC	61
RPOC_EC_2218_2337	CTGGCAGGTATGCGTGGTCTGATG	62	CGCACCGTGGGTTGAGATGAAGTAC	63
RPOC_EC_2218_2337_2	CTTGCTGGTATGCGTGGTCTGATG	64	CGCACCATGCGTAGAGATGAAGTAC	65
RPOC_EC_808_889	CGTCGGGTGATTAACCGTAACAACCG	66	GTTTTTCGTTGCGTACGATGATGTC	67
RPOC_EC_808_891	CGTCGTGTAATTAACCGTAACAACCG	68	ACGTTTTTCGTTTTGAACGATAATGCT	69
RPOC_EC_993_1059	CAAAGGTAAGCAAGGTCGTTTCCGTCA	70	CGAACGGCCTGAGTAGTCAACACG	71
RPOC_EC_993_1059_2	CAAAGGTAAGCAAGGACGTTTCCGTCA	72	CGAACGGCCAGAGTAGTCAACACG	73
TUFB_EC_239_303	TAGACTGCCCAGGACACGCTG	74	GCCGTCCATCTGAGCAGCACC	75
TUFB_EC_239_303_2	TTGACTGCCCAGGTCACGCTG	76	GCCGTCCATTTGAGCAGCACC	77
TUFB_EC_976_1068	AACTACCGTCCGCAGTTCTACTTCC	78	GTTGTCGCCAGGCATAACCATTTC	79
TUFB_EC_976_1068_2	AACTACCGTCCTCAGTTCTACTTCC	80	GTTGTCACCAGGCATTACCATTTC	81
TUFB_EC_985_1062	CCACAGTTCTACTTCCGTACTACTGACG	82	TCCAGGCATTACCATTTCTACTCCTTCTGG	83
RPLB_EC_650_762	GACCTACAGTAAGAGGTTCTGTAATGAACC	84	TCCAAGTGCTGGTTTACCCCATGG	85
RPLB_EC_688_757	CATCCACACGGTGGTGGTGAAGG	86	GTGCTGGTTTACCCCATGGAGT	87
RPOC_EC_1036_1126	CGTGTTGACTATTCGGGGCGTTCAG	88	ATTCAAGAGCCATTTCTTTTGGTAAACCAC	89

RPOB_EC_3762_3865	TCAACAACCTCTTGGAGGTAAAGCTCAGT	90	TTTCTTGAAGAGTATGAGCTGCTCCGTAAG	91
RPLB_EC_688_771	CATCCACACGGTGGTGGTGAAGG	92	TGTTTTGTATCCAAGTGCTGGTTTACCCC	93
VALS_EC_1105_1218	CGTGGCGGCGTGGTTATCGA	94	CGGTACGAACTGGATGTCGCCGTT	95
RPOB_EC_1845_1929	TATCGCTCAGGCGAACTCCAAC	96	GCTGGATTCGCCTTTGCTACG	97
RPLB_EC_669_761	TGTAATGAACCCTAATGACCATCCACACGG	98	CCAAGTGCTGGTTTACCCCATGGAGTA	99
RPLB_EC_671_762	TAATGAACCCTAATGACCATCCACACGGTG	100	TCCAAGTGCTGGTTTACCCCATGGAG	101
RPOB_EC_3775_3858	CTTGGAGGTAAGTCTCATTTTGGTGGGCA	102	CGTATAAGCTGCACCATAAGCTTGTAATGC	103
VALS_EC_1833_1943	CGACGCGCTGCGCTTCAC	104	GCGTTCCACAGCTTGTTGCAGAAG	105
RPOB_EC_1336_1455	GACCACCTCGGCAACCGT	106	TTCGCTCTCGGCCTGGCC	107
TUFB_EC_225_309	GCACTATGCACACGTAGATTGTCCTGG	108	TATAGCACCATCCATCTGAGCGGCAC	109
DNAK_EC_428_522	CGGCGTACTTCAACGACAGCCA	110	CGCGGTCGGCTCGTTGATGA	111
VALS_EC_1920_1970	CTTCTGCAACAAGCTGTGGAACGC	112	TCGCAGTTCATCAGCACGAAGCG	113
TUFB_EC_757_867	AAGACGACCTGCACGGGC	114	GCGCTCCACGTCTTCACGC	115
23S_EC_2646_2765	CTGTTCTTAGTACGAGAGGACC	116	TTCGTGCTTAGATGCTTTCAG	117
16S_EC_969_1078_3P	ACGCGAAGAACCTTACpC	118	ACGACACGAGCpTpGACGAC	119
16S_EC_972_1075_4P	CGAAGAACpCpTTACC	120	ACACGAGCpTpGAC	121
16S_EC_972_1075	CGAAGAACCTTACC	122	ACACGAGCTGAC	123
23S_EC_-347_59	CCTGATAAGGGTGAGGTCG	124	ACGTCCTTCATCGCCTCTGA	125
23S_EC_-7_450	GTTGTGAGGTTAAGCGACTAAG	126	CTATCGGTCAGTCAGGAGTAT	127
23S_EC_-7_910	GTTGTGAGGTTAAGCGACTAAG	128	TTGCATCGGGTTGGTAAGTC	129
23S_EC_430_1442	ATACTCCTGACTGACCGATAG	130	AACATAGCCTTCTCCGTCC	131
23S_EC_891_1931	GACTTACCAACCCGATGCAA	132	TACCTTAGGACCGTTATAGTTACG	133
23S_EC_1424_2494	GGACGGAGAAGGCTATGTT	134	CCAAACACCGCCGTCGATAT	135
23S_EC_1908_2852	CGTAACTATAACGGTCCTAAGGTA	136	GCTTACACACCCGGCCTATC	137
23S_EC_2475_3209	ATATCGACGGCGGTGTTTGG	138	GCGTGACAGGCAGGTATTC	139
16S_EC_-60_525	AGTCTCAAGAGTGAACACGTAA	140	GCTGCTGGCACGGAGTTA	141
16S_EC_326_1058	GACACGGTCCAGACTCCTAC	142	CCATGCAGCACCTGTCTC	143
16S_EC_705	GATCTGGAGGAATACCGGTG	144	ACGGTTACCTTGTTACGACT	145

_1512				
16S_EC_1268_1775	GAGAGCAAGCGGACCTCATA	146	CCTCCTGCGTGCAAAGC	147
GROL_EC_941_1060	TGGAAGATCTGGGTCAGGC	148	CAATCTGCTGACGGATCTGAGC	149
INFB_EC_1103_1191	GTCGTGAAAACGAGCTGGAAGA	150	CATGATGGTCACAACCGG	151
HFLB_EC_1082_1168	TGGCGAACCTGGTGAACGAAGC	152	CTTTCGCTTTCTCGAACTCAACCAT	153
INFB_EC_1969_2058	CGTCAGGGTAAATTCCGTGAAGTTAA	154	AACTTCGCCTTCGGTCATGTT	155
GROL_EC_219_350	GGTGAAAGAAGTTGCCTCTAAAGC	156	TTCAGGTCCATCGGGTTCATGCC	157
VALS_EC_1105_1214	CGTGGCGGCGTGGTTATCGA	158	ACGAACTGGATGTCGCCGTT	159
16S_EC_556_700	CGGAATTACTGGGCGTAAAG	160	CGCATTTCACCGCTACAC	161
RPOC_EC_1256_1315	ACCCAGTGCTGCTGAACCGTGC	162	GTTCAAATGCCTGGATACCCA	163
16S_EC_774_894	GGGAGCAAACAGGATTAGATAC	164	CGTACTCCCCAGGCG	165
RPOC_EC_1584_1643	TGGCCCGAAAGAAGCTGAGCG	166	ACGCGGGCATGCAGAGATGCC	167
16S_EC_1082_1196	ATGTTGGGTTAAGTCCCGC	168	TGACGTCATCCCCACCTTCC	169
16S_EC_1389_1541	CTTGTACACACCGCCCGTC	170	AAGGAGGTGATCCAGCC	171
16S_EC_1303_1407	CGGATTGGAGTCTGCAACTCG	172	GACGGGCGGTGTGTACAAG	173
23S_EC_23_130	GGTGGATGCCTTGGC	174	GGGTTTCCCCATTCGG	175
23S_EC_187_256	GGGAACTGAAACATCTAAGTA	176	TTCGCTCGCCGCTAC	177
23S_EC_1602_1703	TACCCCAAACCGACACAGG	178	CCTTCTCCCGAAGTTACG	179
23S_EC_1685_1842	CCGTAACTTCGGGAGAAGG	180	CACCGGGCAGGCGTC	181
23S_EC_1827_1949	GACGCCTGCCCGGTGC	182	CCGACAAGGAATTTCGCTACC	183
23S_EC_2434_2511	AAGGTACTCCGGGGATAACAGGC	184	AGCCGACATCGAGGTGCCAAAC	185
23S_EC_2599_2669	GACAGTTCGGTCCCTATC	186	CCGGTCCTCTCGTACTA	187
23S_EC_2653_2758	TAGTACGAGAGGACCGG	188	TTAGATGCTTTCAGCACTTATC	189
23S_BS_-68_21	AAACTAGATAACAGTAGACATCAC	190	GTGCGCCCTTTCTAACTT	191
16S_EC_8_358	AGAGTTTGATCATGGCTCAG	192	ACTGCTGCCTCCCGTAG	193
16S_EC_314_575	CACTGGAACTGAGACACGG	194	CTTTACGCCCAGTAATTCCG	195
16S_EC_518_795	CCAGCAGCCGCGGTAATAC	196	GTATCTAATCCTGTTTGCTCCC	197
16S_EC_683_985	GTGTAGCGGTGAAATGCG	198	GGTAAGGTTCTTCGCGTTG	199
16S_EC_937_1240	AAGCGGTGGAGCATGTGG	200	ATTGTAGCACGTGTGTAGCCC	201

16S_EC_1195_1541	CAAGTCATCATGGCCCTTA	202	AAGGAGGTGATCCAGCC	203
16S_EC_8_1541	AGAGTTTGATCATGGCTCAG	204	AAGGAGGTGATCCAGCC	205
23S_EC_1831_1936	ACCTGCCCAGTGCTGGAAG	206	TCGCTACCTTAGGACCGT	207
16S_EC_1387_1513	GCCTTGTACACACCTCCCGTC	208	CACGGCTACCTTGTTACGAC	209
16S_EC_1390_1505	TTGTACACACCGCCCGTCATAC	210	CCTTGTTACGACTTCACCCC	211
16S_EC_1367_1506	TACGGTGAATACGTTCCCGGG	212	ACCTTGTTACGACTTCACCCCA	213
16S_EC_804_929	ACCACGCCGTAAACGATGA	214	CCCCCGTCAATTCCTTTGAGT	215
16S_EC_791_904	GATACCCTGGTAGTCCACACCG	216	GCCTTGCGACCGTACTCCC	217
16S_EC_789_899	TAGATACCCTGGTAGTCCACGC	218	GCGACCGTACTCCCCAGG	219
16S_EC_1092_1195	TAGTCCCGCAACGAGCGC	220	GACGTCATCCCCACCTTCCTCC	221
23S_EC_2586_2677	TAGAACGTCGCGAGACAGTTCG	222	AGTCCATCCCGGTCCTCTCG	223
HEXAMER_EC_61_362	GAGGAAAGTCCGGGCTC	224	ATAAGCCGGGTTCTGTCG	225
RNASEP_BS_43_384	GAGGAAAGTCCATGCTCGC	226	GTAAGCCATGTTTTGTTCCATC	227
RNASEP_EC_61_362	GAGGAAAGTCCGGGCTC	228	ATAAGCCGGGTTCTGTCG	229
YAED_TRNA_ALA-RRNH_EC_513_49	GCGGGATCCTCTAGAGGTGTTAAATAGCCTGGCAG	230	GCGGGATCCTCTAGAAGACCTCCTGCGTGCAAAGC	231
RNASEP_SA_31_379	GAGGAAAGTCCATGCTCAC	232	ATAAGCCATGTTCTGTTCCATC	233
16S_EC_1082_1541	ATGTTGGGTTAAGTCCCGC	234	AAGGAGGTGATCCAGCC	235
16S_EC_556_795	CGGAATTACTGGGCGTAAAG	236	GTATCTAATCCTGTTTGCTCCC	237
16S_EC_1082_1196_10G	ATGTTGGGTTAAGTCCCGC	238	TGACGTCATGCCCACCTTCC	239
16S_EC_1082_1196_10G_11G	ATGTTGGGTTAAGTCCCGC	240	TGACGTCATGGCCACCTTCC	241
TRNA_ILERRNH_ASPRRNH_EC_32_41	GCGGGATCCTCTAGACCTGATAAGGGTGAGGTCG	242	GCGGGATCCTCTAGAGCGTGACAGGCAGGTATTC	243
16S_EC_969_1407	ACGCGAAGAACCTTACC	244	GACGGGCGGTGTGTACAAG	245
16S_EC_683_1323	GTGTAGCGGTGAAATGCG	246	CGAGTTGCAGACTGCGATCCG	247
16S_EC_49_894	TAACACATGCAAGTCGAACG	248	CGTACTCCCCAGGCG	249
16S_EC_49_1078	TAACACATGCAAGTCGAACG	250	ACGACACGAGCTGACGAC	251
CYA_BA_1349_1447	ACAACGAAGTACAATACAAGAC	252	CTTCTACATTTTTAGCCATCAC	253

16S_EC_1090_1196_2	TTAAGTCCCGCAACGAGCGCAA	254	TGACGTCATCCCCACCTTCCTC	255
16S_EC_405_527	TGAGTGATGAAGGCCTTAGGGTTGTAAA	256	CGGCTGCTGGCACGAAGTTAG	257
GROL_EC_496_596	ATGGACAAGGTTGGCAAGGAAGG	258	TAGCCGCGGTCGAATTGCAT	259
GROL_EC_511_593	AAGGAAGGCGTGATCACCGTTGAAGA	260	CCGCGGTCGAATTGCATGCCTTC	261
VALS_EC_1835_1928	ACGCGCTGCGCTTCAC	262	TTGCAGAAGTTGCGGTAGCC	263
RPOB_EC_1334_1478	TCGACCACCTGGGCAACC	264	ATCAGGTCGTGCGGCATCA	265
DNAK_EC_420_521	CACGGTGCCGGCGTACT	266	GCGGTCGGCTCGTTGATGAT	267
RPOB_EC_3776_3853	TTGGAGGTAAGTCTCATTTTGGTGG	268	AAGCTGCACCATAAGCTTGTAATGC	269
RPOB_EC_3802_3885	CAGCGTTTCGGCGAAATGGA	270	CGACTTGACGGTTAACATTTCCTG	271
RPOB_EC_3799_3888	GGGCAGCGTTTCGGCGAAATGGA	272	GTCCGACTTGACGGTCAACATTTCCTG	273
RPOC_EC_2146_2245	CAGGAGTCGTTCAACTCGATCTACATGAT	274	ACGCCATCAGGCCACGCAT	275
ASPS_EC_405_538	GCACAACCTGCGGCTGCG	276	ACGGCACGAGGTAGTCGC	277
RPOC_EC_1374_1455	CGCCGACTTCGACGGTGACC	278	GAGCATCAGCGTGCGTGCT	279
TUFB_EC_957_1058	CCACACGCCGTTCTTCAACAACT	280	GGCATCACCATTTCCTTGTCCTTCG	281
16S_EC_7_122	GAGAGTTTGATCCTGGCTCAGAACGAA	282	TGTTACTCACCCGTCTGCCACT	283
VALS_EC_610_727	ACCGAGCAAGGAGACCAGC	284	TATAACGCACATCGTCAGGGTGA	285

For evaluation in the laboratory, five species of bacteria were selected including three γ-proteobacteria (*E. coli, K. pneumoniae* and *P. auergiosa*) and two low G+C gram positive bacteria (*B. subtilitis* and *S. aureus*). The identities of the organisms were not revealed to the laboratory technicians.

Bacteria were grown in culture, DNA was isolated and processed, and PCR performed using standard protocols. Following PCR, all samples were desalted, concentrated, and analyzed by Fourier Transform Ion Cyclotron Resonance (FTICR) mass spectrometry. Due to the extremely high precision of the FTICR, masses could be measured to within 1 Da and unambiguously deconvoluted to a single base composition. The measured base compositions were compared with the known base composition signatures in our database. As expected when using broad range survey 16S primers, several phylogenetic near-neighbor organisms were difficult to distinguish from our test organisms. Additional non-ribosomal primers were used to triangulate and further resolve these clusters.

An example of the use of primers directed to regions of RNA polymerase B (rpoB) is shown in Figure 19. This gene has the potential to provide broad priming and resolving capabilities. A pair of primers directed against a conserved region of rpoB provided distinct base composition signatures that helped resolve the tight enterobacteriae cluster. Joint probability estimates of the signatures from each of the primers resulted in the identification of a single organism that matched the identity of the test sample. Therefore a combination of a small number of primers that amplify selected regions of the 16S ribosomal RNA gene and a few additional primers that amplify selected regions of protein encoding genes provide sufficient information to detect and identify all bacterial pathogens.

Example 16: Detection of *Staphylococcus aureus* in Blood Samples

Blood samples in an analysis plate were spiked with genomic DNA equivalent of 10^3 organisms/ml of *Staphylococcus aureus*. A single set of 16S rRNA primers was used for amplification. Following PCR, all samples were desalted, concentrated, and analyzed by Fourier Transform Ion Cyclotron Resonance (FTICR) mass spectrometry. In each of the spiked wells, strong signals were detected which are consistent with the expected BCS of the *S. aureus* amplicon (Figure 20). Furthermore, there was no robotic carryover or contamination in any of the blood only or water blank wells. Methods similar to this one will be applied for other clinically relevant samples including, but not limited to: urine and throat or nasal swabs.

Example 17: Detection and Serotyping of Viruses

The virus detection capability of the present invention was demonstrated in collaboration with Naval health officers using adenoviruses as an example.

All available genomic sequences for human adenoviruses available in public databases were surveyed. The hexon gene was identified as a candidate likely to have broad specificity across all serotypes. Four primer pairs were selected from a group of primers designed to yield broad coverage across the majority of the adenoviral strain types (Table 9) wherein Tp = 5'propynylated uridine and Cp = 5'propynylated cytidine.

Table 9: Intelligent Primer Pairs for Serotyping of Adenoviruses

Primer Pair Name	Forward Primer Sequence	Forward SEQ ID NO:	Reverse Primer Sequence	Reverse SEQ ID NO:
HEX_HAD7+4+2	AGACCCAATTACATTGGCTT	286	CCAGTGCTGTTGTAGTACAT	287

1_934_995				
HEX_HAD7+4+21_976_1050	ATGTACTACAACAGTACTGG	288	CAAGTCAACCACAGCATTCA	289
HEX_HAD7+4+21_970_1059	GGGCTTATGTACTACAACAG	290	TCTGTCTTGCAAGTCAACCAC	291
HEX_HAD7+3_771_827	GGAATTTTTTGATGGTAGAGA	292	TAAAGCACAATTTCAGGCG	293
HEX_HAD4+16_746_848	TAGATCTGGCTTTCTTTGAC	294	ATATGAGTATCTGGAGTCTGC	295
HEX_HAD7_509_578	GGAAAGACATTACTGCAGACA	296	CCAACTTGAGGCTCTGGCTG	297
HEX_HAD4_1216_1289	ACAGACACTTACCAGGGTG	298	ACTGTGGTGTCATCTTTGTC	299
HEX_HAD21_515_567	TCACTAAAGACAAAGGTCTTCC	300	GGCTTCGCCGTCTGTAATTTC	301
HEX_HAD_1342_1469	CGGATCCAAGCTAATCTTTGG	302	GGTATGTACTCATAGGTGTTGGTG	303
HEX_HAD7+4+21_934_995P	AGACpCpCAATTpACpATpTGGCTT	304	CpCpAGTGCTGTpTpGTAGTACAT	305
HEX_HAD7+4+21_976_1050P	ATpGTpACTpACAACAGTACpTpGG	306	CAAGTpCpAACCACAGCATpTpCA	307
HEX_HAD7+4+21_970_1059P	GGGCpTpTATpGTpACTACAACpAG	308	TCTGTpCpTTGCAAGTpCpAACCAC	309
HEX_HAD7+3_771_827P	GGAATTpTpTpTpTGATGGTAGAGA	310	TAAAGCACAATpTpTpCpAGGCG	311
HEX_HAD4+16_746_848P	TAGATCTGGCTpTpTpCpTTTGAC	312	ATATGAGTATpCpTpGGAGTpCpTGC	313
HEX_HAD_1342_1469P	CGGATpCCAAGCpTAATCpTpTTGG	314	GGTATGTACTCATAGGTGTpTpGGTG	315
HEX_HAD7+21+3_931_1645	AACAGACCCAATTACATTGGCTT	316	GAGGCACTTGTATGTGGAAAGG	317
HEX_HAD4+2_925_1469	ATGCCTAACAGACCCAATTACAT	318	TTCATGTAGTCGTAGGTGTTGG	319
HEX_HAD7+21+3_384_953	CGCGCCTAATACATCTCAGTGGAT	320	AAGCCAATGTAATTGGGTCTGTT	321
HEX_HAD4+2_345_947	CTACTCTGGCACTGCCTACAAC	322	ATGTAATTGGGTCTGTTAGGCAT	323
HEX_HAD2_772_865	CAATCCGTTCTGGTTCCGGATGAA	324	CTTGCCGGTCGTTCAAAGAGGTAG	325
HEX_HAD7+4+21_73_179	AGTCCGGGTCTGGTGCAG	326	CGGTCGGTGGTCACATC	327
HEX_HAD7+4+21_1_54	ATGGCCACCCCATCGATG	328	CTGTCCGGCGATGTGCATG	329
HEX_HAD7+4+21_1612_1718	GGTCGTTATGTGCCTTTCCACAT	330	TCCTTTCTGAAGTTCCACTCATAGG	331
HEX_HAD7+4+21_2276_2368	ACAACATTGGCTACCAGGGCTT	332	CCTGCCTGCTCATAGGCTGGAAGTT	333

These primers also served to clearly distinguish those strains responsible for most disease (types 3, 4, 7 and 21) from all others. DNA isolated from field samples known to contain adenoviruses were tested using the hexon gene PCR primers, which provided unambiguous strain identification for all samples. A single sample was found to contain a mixture of two viral DNAs belonging to strains 7 and 21.

Test results (Figure 21) showed perfect concordance between predicted and observed base composition signatures for each of these samples. Classical serotyping results confirmed each of these observations. Processing of viral samples directly from collection material such as throat swabs rather than from isolated DNA, will result in a significant increase in throughput, eliminating the need for virus culture.

Example 18: Broad Rapid Detection and Strain Typing of Respiratory Pathogens for Epidemic Surveillance

Genome preparation: Genomic materials from culture samples or swabs were prepared using a modified robotic protocol using DNeasy™ 96 Tissue Kit, Qiagen). Cultures of *Streptococcus pyogenes* were pelleted and transferred to a 1.5 mL tube containing 0.45 g of 0.7 mm Zirconia beads (Biospec Products, Inc.). Cells were lysed by shaking for 10 minutes at a speed of 19 1/s using a MM300 Vibration Mill (Retsch, Germany). The samples were centrifuged for 5 min and the supernatants transferred to deep well blocks and processed using the manufacture's protocol and a Qiagen 8000 BioRobot.

PCR: PCR reactions were assembled using a Packard MPII liquid handling platform and were performed in 50 µL volume using 1.8 units each of Platinum Taq (Invitrogen) and Hotstart PFU Turbo (Stratagene) polymerases. Cycling was performed on a DNA Engine Dyad (MJ Research) with cycling conditions consisting of an initial 2 min at 95°C followed by 45 cycles of 20 s at 95°C, 15 s at 58°C, and 15 s at 72°C.

Broad-range primers: PCR primer design for base composition analysis from precise mass measurements is constrained by an upper limit where ionization and accurate deconvolution can be achieved. Currently, this limit is approximately 140 base pairs. Primers designed to broadly conserved regions of bacterial ribosomal RNAs (16 and 23S) and the gene encoding ribosomal protein L3 (rpoC) are shown in Table 10.

Table 10: Broad Range Primer Pairs

Target Gene	Direction	Primer	SEQ ID NO	Length of Amplicon
16S_1	F	GGATTAGAGACCCTGGTAGTCC	334	116
16S_1	R	GGCCGTACTCCCCAGGCG	335	116
16S_2	F	TTCGATGCAACGCGAAGAACCT	336	115
16S_2	R	ACGAGCTGACGACAGCCATG	337	115
23S	F	TCTGTCCCTAGTACGAGAGGACCGG	338	118
23S	R	TGCTTAGATGCTTTCAGC	339	118

rpoC	F	CTGGCAGGTATGCGTGGTCTGATG	340	121
rpoC	R	CGCACCGTGGGTTGAGATGAAGTAC	341	121

Emm-typing primers: The allelic profile of a GAS strain by Multilocus Sequencing Technique (MLST) can be obtained by sequencing the internal fragments of seven housekeeping genes. The nucleotide sequences for each of these housekeeping genes, for 212 isolates of GAS (78 distinct emm types), are available (www.mlst.net). This corresponds to one hundred different allelic profiles or unique sequence types, referred to by Enright et al. as ST1-ST100 (Enright, M. C., *et al.*, *Infection and Immunity* **2001,** *69*, 2416-2427). For each sequence type, we created a virtual transcript by concatenating sequences appropriate to their allelic profile from each of the seven genes. MLST primers were designed using these sequences and were constrained to be within each gene loci. Twenty-four primer pairs were initially designed and tested against the sequenced GAS strain 700294. A final subset of six primer pairs Table 11 was chosen based on a theoretical calculation of minimal number of primer pairs that maximized resolution of between emm types.

Table 11: Drill-Down Primer Pairs Used in Determining emm-type

Target Gene	Direction	Primer	SEQ ID NO	Length of Amplicon
gki	F	GGGGATTCAGCCATCAAAGCAGCTATTGAC	342	116
gki	R	CCAACCTTTTCCACAACAGAATCAGC	343	116
gtr	F	CCTTACTTCGAACTATGAATCTTTTGGAAG	344	115
gtr	R	CCCATTTTTTCACGCATGCTGAAAATATC	345	115
murI	F	CGCAAAAAAATCCAGCTATTAGC	346	118
murI	R	AAACTATTTTTTTAGCTATACTCGAACAC	347	118
mutS	F	ATGATTACAATTCAAGAAGGTCGTCACGC	348	121
mutS	R	TTGGACCTGTAATCAGCTGAATACTGG	349	121
xpt	F	GATGACTTTTTAGCTAATGGTCAGGCAGC	350	122
xpt	R	AATCGACGACCATCTTGGAAAGATTTCTC	351	122
yqiL	F	GCTTCAGGAATCAATGATGGAGCAG	352	119
yqiL	R	GGGTCTACACCTGCACTTGCATAAC	353	119

Microbiology: GAS isolates were identified from swabs on the basis of colony morphology and beta-hemolysis on blood agar plates, gram stain characteristics, susceptibility to bacitracin, and positive latex agglutination reactivity with group A-specific antiserum.

Sequencing: Bacterial genomic DNA samples of all isolates were extracted from freshly grown GAS strains by using QIAamp DNA Blood Mini Kit (Qiagen, Valencia, CA) according to the procedures described by the manufacture. Group A streptococcal cells were subjected to PCR and sequence analysis using emm-gene specific PCR as previously described (Beall, B., *et al. J. Clin. Micro.*, **1996**, *34*, 953-958; Facklam, R., *et al. Emerg. Infect. Dis.* **1999**, *5*, 247-253). Homology searches on DNA sequences were conducted against known emm sequences present in (www.cdc.gov/ncidod/biotech/infotech_hp.html). For MLST analysis, internal fragments of seven housekeeping genes, were amplified by PCR and analyzed as previously described (Enright, M. C., *et al.*, *Infection and Immunity* **2001,** *69*, 2416-2427). The emm-type was determined from comparison to the MLST database.

Broad Range Survey/Drill-Down Process (100): For *Streptococcus pyogenes*, the objective was the identification of a signature of the virulent epidemic strain and determination of its emm-type. Emm-type information is useful both for treatment considerations and epidemic surveillance. A total of 51 throat swabs were taken both from healthy recruits and from hospitalized patients in December 2002, during the peak of a GAS outbreak at a military training camp. Twenty-seven additional isolates from previous infections ascribed to GAS were also examined. Initially, isolated colonies were examined both from throat culture samples and throat swabs directly without the culture step. The latter path can be completed within 6-12 hours providing information on a significant number of samples rapidly enough to be useful in managing an ongoing epidemic.

The process of broad range survey/drill-down **(200)** is shown in Figure 22. A clinical sample such as a throat swab is first obtained from an individual **(201)**. Broad range survey primers are used to obtain amplification products from the clinical sample **(202)** which are analyzed to determine a BCS **(203)** from which a species is identified **(204)**. Drill-down primers are then employed to obtain PCR products **(205)** from which specific information is obtained about the species (such as Emm-type) **(206)**.

Broad Range Survey Priming: Genomic regions targeted by the broad range survey primers were selected for their ability to allow amplification of virtually all known species of bacteria and for their capability to distinguish bacterial species from each other by base composition analysis. Initially, four broad-range PCR target sites were selected and the primers were synthesized and tested. The targets included universally conserved regions of 16S and 23S rRNA, and the gene encoding ribosomal protein L3 (rpoC).

While there was no special consideration of *Streptococcus pyogenes* in the selection of the broad range survey primers (which were optimized for distinguishing all important pathogens from each other), analysis of genomic sequences showed that the base compositions of these regions distinguished *Streptococcus pyogenes* from other respiratory pathogens and normal flora, including closely related species of *streptococci*, *staphylococci*, and *bacilli* (Figure 23).

Drill Down Priming (Emm-Typing): In order to obtain strain-specific information about the epidemic, a strategy was designed to measure the base compositions of a set of fast clock target genes to generate strain-specific signatures and simultaneously correlate with emm-types. In classic MLST analysis, internal fragments of seven housekeeping genes (*gki, gtr, murI, mutS, recP, xpt, yqiL*) are amplified, sequenced and compared to a database of previously studied isolates whose emm-types have been determined (Horner, M. J. ,*et al. Fundamental and Applied Toxicology*, **1997,** *36*, 147). Since the analysis enabled by the present embodiment of the present invention provides base composition data rather than sequence data, the challenge was to identify the target regions that provide the highest resolution of species and least ambiguous emm-classification. The data set from Table 2 of Enright et al. (Enright, M. C., *et al. Infection and Immunity*, **2001,** *69*, 2416-2427) to bioinformatically construct an alignment of concatenated alleles of the seven housekeeping genes from each of 212 previously emm-typed strains, of which 101 were unique sequences that represented 75 distinct emm-types. This alignment was then analyzed to determine the number and location of the optimal primer pairs that would maximize strain discrimination strictly on base composition data.

An example of assignment of BCSs of PCR products is shown in Figure 24 where PCR products obtained using the *gtr* primer (a drill-down emm-typing primer) from two different swab samples were analyzed (sample 12 – top and sample 10 – bottom). The deconvoluted ESI-FCTIR spectra provide accurate mass measurements of both strands of the PCR products, from which a series of candidate BCSs were calculated from the measured mass (and within the measured mass uncertainty). The identification of complementary candidate BCSs from each strand provides a means for unambiguous assignment of the BCS of the PCR product. BCSs and molecular masses for each strand of the PCR product from the two different samples are also shown in Figure 24. In this case, the determination of BCSs for the two samples resulted in the identification of the emm-type of *Streptococcus pyogenes* – sample 12 was identified as emm-type 3 and sample 10 was identified as emm-type 6.

The results of the composition analysis using the six primer pairs, 5'-emm gene sequencing and MLST gene sequencing method for the GAS epidemic at a military training facility are compared in Figure 25. The base composition results for the six primer pairs showed a perfect concordance with 5'-emm gene sequencing and MLST sequencing methods. Of the 51 samples taken during the peak of the epidemic, all but three had identical compositions and corresponded to emm-type 3. The three outliers, all from healthy individuals, probably represent non-epidemic strains harbored by asymptomatic carriers. Samples 52-80, which were archived from previous infections from Marines at other naval training facilities, showed a much greater heterogeneity of composition signatures and emm-types.

Example 19: Base Composition Probability Clouds

Figure 18 illustrates the concept of base composition probability clouds via a pseudo-four dimensional plot of base compositions of enterobacteria including *Y. pestis, Y. psuedotuberculosis, S. typhimurium, S. typhi, Y. enterocolitica, E. coli* K12, and *E. coli* O157:H7. In the plot of Figure 18, A, C and G compositions correspond to the x, y and z axes respectively whereas T compositions are represented by the size of the sphere at the junction of the x, y and z coordinates. There is no absolute requirement for having a particular nucleobase composition associated with a particular axis. For example, a plot could be designed wherein G, T and C compositions correspond to the x, y and z axes respectively whereas the A composition corresponds to the size of the sphere at the junction of the x, y and z coordinates. Furthermore, a different representation can be made of the "pseudo fourth" dimension i.e.: other than the size of the sphere at junction of the x, y and z coordinates. For example, a symbol having vector information such as an arrow or a cone can be rotated at an angle which varies proportionally with the composition of the nucleobase corresponding to the pseudo fourth dimension. The choice of axes and pseudo fourth dimensional representation is typically made with the aim of optimal visualization of the data being presented.

A similar base composition probability cloud analysis has been presented for a series of viruses in U.S. provisional patent application Serial No. 60/431,319, which is commonly owned and incorporated herein by reference in its entirety. In this base composition probability cloud analysis, the closely related Dengue virus types 1-4 are clearly distinguishable from each other. This example is indicative of a challenging scenario for

species identification based on BCS analysis because RNA viruses have a high mutation rate, it would be expected to be difficult to resolve closely related species. However, as this example illustrates, BCS analysis, aided by base composition probability cloud analysis is capable of resolution of closely related viral species.

A base composition probability cloud can also be represented as a three dimensional plot instead of a pseudo-four dimensional plot. An example of such a three dimensional plot is a plot of G, A and C compositions correspond to the x, y and z axes respectively, while the composition of T is left out of the plot. Another such example is a plot where the compositions of all four nucleobases is included: G, A and C+T compositions correspond to the x, y and z axes respectively. As for the pseudo-four dimensional plots, the choice of axes for a three dimensional plot is typically made with the aim of optimal visualization of the data being presented.

Example 20: Biochemical Processing of Large Amplification Products for Analysis by Mass Spectrometry

In the example illustrated in Figure 26, a primer pair which amplifies a 986 bp region of the 16S ribosomal gene in E. coli (K12) was digested with a mixture of 4 restriction enzymes: *BstN1, BsmF1, Bfa1, and Nco1*. Figure 26(a) illustrates the complexity of the resulting ESI-FTICR mass spectrum which contains multiple charge states of multiple restriction fragments. Upon mass deconvolution to neutral mass, the spectrum is significantly simplified and discrete oligonucleotide pairs are evident (Figure 26b). When base compositions are derived from the masses of the restriction fragments, perfect agreement is observed for the known sequence of nucleotides 1-856 (Figure 26c); the batch of *Nco1* enzyme used in this experiment was inactive and resulted in a missed cleavage site and a 197-mer fragment went undetected as it is outside the mass range of the mass spectrometer under the conditions employed. Interestingly however, both a forward and reverse strand were detected for each fragment measured (solid and dotted lines in, respectively) within 2 ppm of the predicted molecular weights resulting in unambiguous determination of the base composition of 788 nucleotides of the 985 nucleotides in the amplicon. The coverage map offers redundant coverage as both 5' to 3' and 3' to 5' fragments are detected for fragments covering the first 856 nucleotides of the amplicon.

This approach is in many ways analogous to those widely used in MS-based proteomics studies in which large intact proteins are digested with trypsin, or other

proteolytic enzyme(s), and the identity of the protein is derived by comparing the measured masses of the tryptic peptides with theoretical digests. A unique feature of this approach is that the precise mass measurements of the complementary strands of each digest product allow one to derive a de novo base composition for each fragment, which can in turn be "stitched together" to derive a complete base composition for the larger amplicon. An important distinction between this approach and a gel-based restriction mapping strategy is that, in addition to determination of the length of each fragment, an unambiguous base composition of each restriction fragment is derived. Thus, a single base substitution within a fragment (which would not be resolved on a gel) is readily observed using this approach. Because this study was performed on a 7 Tesla ESI-FTICR mass spectrometer, better than 2 ppm mass measurement accuracy was obtained for all fragments. Interestingly, calculation of the mass measurement accuracy required to derive unambiguous base compositions from the complementary fragments indicates that the highest mass measurement accuracy actually required is only 15 ppm for the 139 bp fragment (nucleotides 525-663). Most of the fragments were in the 50-70 bp size-range which would require mass accuracy of only ~50 ppm for unambiguous base composition determination. This level of performance is achievable on other more compact, less expensive MS platforms such as the ESI-TOF suggesting that the methods developed here could be widely deployed in a variety of diagnostic and human forensic arenas.

This example illustrates an alternative approach to derive base compositions from larger PCR products. Because the amplicons of interest cover many strain variants, for some of which complete sequences are not known, each amplicon can be digested under several different enzymatic conditions to ensure that a diagnostically informative region of the amplicon is not obscured by a "blind spot" which arises from a mutation in a restriction site. The extent of redundancy required to confidently map the base composition of amplicons from different markers, and determine which set of restriction enzymes should be employed and how they are most effectively used as mixtures can be determined. These parameters will be dictated by the extent to which the area of interest is conserved across the amplified region, the compatibility of the various restriction enzymes with respect to digestion protocol (buffer, temperature, time) and the degree of coverage required to discriminate one amplicon from another.

Example 21: Identification of members of the Viral Genus *Orthopoxvirus*

Primer sites were identified on three essential viral genes – the DNA-dependent polymerase (DdDp), and two sub-units of DNA-dependent RNA polymerases A and B (DdRpA and DdRpB). These intelligent primers designed to identify members of the viral genus *Orthopoxvirus* are shown in Table 12 wherein Tp = 5'propynylated uridine and Cp = 5'propynylated cytidine.

Table 12: Intelligent Primer Pairs for Identification of members of the Viral Genus *Orthopoxvirus*

Primer Pair Name	Forward Primer Sequence	Forward SEQ ID NO:	Reverse Primer Sequence	Reverse SEQ ID NO:
A25L_NC00161 1_28_127	GTACTGAATCCGCCTAAG	354	GTGAATAAAGTATCGCCCTAA TA	355
A18R_NC00161 1_100_207	GAAGTTGAACCGGGATCA	356	ATTATCGGTCGTTGTTAATGT	357
A18R_NC00161 1_1348_1445	CTGTCTGTAGATAAACTAGGAT T	358	CGTTCTTCTCTGGAGGAT	359
E9L_NC001611 1119_1222	CGATACTACGGACGC	360	CTTTATGAATTACTTTACATA T	361
K8R_NC001611 221_311	CTCCTCCATCACTAGGAA	362	CTATAACATTCAAAGCTTATT G	363
A24R_NC00161 1_795_878	CGCGATAATAGATAGTGCTAAA C	364	GCTTCCACCAGGTCATTAA	365
A25L_NC00161 1_28_127P	GTACpTpGAATpCpCpGCpCpT AAG	366	GTGAATAAAGTATpCpGCpCp CpTpAATA	367
A18R_NC00161 1_100_207P	GAAGTpTpGAACpCpGGGATCA	368	ATTATCGGTpCpGTpTpGTpT pAATGT	369
A18R_NC00161 1_1348_1445P	CTGTpCpTpGTAGATAAACpTp AGGATT	370	CGTTCpTpTpCpTpCpTpGGA GGAT	371
E9L_NC001611 1119_1222P	CGATACpTpACpGGACGC	372	CTTTATGAATpTpACpTpTpT pACATAT	373
K8R_NC001611 221_311P	CTpCpCpTCpCpATCACpTpAG GAA	374	CTATAACATpTpCpAAAGCpT pTpATTG	375
A24R_NC00161 1_795_878P	CGCGATpAATpAGATAGTpGCp TpAAAC	376	GCTTCpCpACpCAGGTpCATp TAA	377

As illustrated in Figure 27, members of the *Orthopoxvirus* genus group can be identified, distinguished from one another, and distinguished from other members of the Poxvirus family using a single pair of primers designed against the DdRpB gene.

Since the primers were designed across regions of high conservation within this genus, the likelihood of missed detection due to sequence variations at these sites is minimized. Further, none of the primers is expected to amplify other viruses or any other DNA, based on the data available in GenBank. This method can be used for all families of viral threat agents and is not limited to members of the *Orthopoxvirus* genus.

Example 22: Identification of Viruses that Cause Viral Hemorrhagic Fevers

In accordance with the present invention an approach of broad PCR priming across several different viral species is employed using conserved regions in the various viral genomes, amplifying a small, yet highly informative region in these organisms, and then analyzing the resultant amplicons with mass spectrometry and data analysis. These regions will be tested with live agents, or with genomic constructs thereof.

Detection of RNA viruses will necessitate a reverse transcription (RT) step prior to the PCR amplification of the TIGER reporter amplicon. To maximize throughput and yield while minimizing the handling of the samples, commercial one-step reverse transcription polymerase chain reaction (RT-PCR) kits will be evaluated for use. If necessary, a one-step RT-PCR mix using our selected DNA polymerase for the PCR portion of the reaction will be developed. To assure there is no variation in our reagent performance all new lots of enzymes, nucleotides and buffers will be individually tested prior to use.

Various modifications of the invention, in addition to those described herein, will be apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims. Each reference cited in the present application is incorporated herein by reference in its entirety